

# On the Achievable Rate of Stationary Rayleigh Flat-Fading Channels with Gaussian Inputs

Meik Dörpinghaus, *Member, IEEE*, and Heinrich Meyr, *Life Fellow, IEEE*

## Abstract

In this work, we consider a discrete-time stationary Rayleigh flat-fading channel with unknown channel state information at transmitter and receiver side. The law of the channel is presumed to be known to the receiver. In addition, we assume the power spectral density of the fading process to be compactly supported. For independent identically distributed (i.i.d.) zero-mean proper Gaussian input distributions, we investigate the achievable rate. One of the main contributions of the present paper is the derivation of two new upper bounds on the achievable rate with zero-mean proper Gaussian input symbols. The first one holds only for the special case of a rectangular power spectral density and depends on the SNR and the spread of the power spectral density. Together with a lower bound on the achievable rate, which is achievable with i.i.d. zero-mean proper Gaussian input symbols, we have found a set of bounds which is tight in the sense that their difference is bounded. Furthermore, we show that the high SNR slope is characterized by a *pre-log* of  $1 - 2f_d$ , where  $f_d$  is the normalized maximum Doppler frequency. This pre-log is equal to the high SNR pre-log of the peak power constrained capacity. Furthermore, we derive an alternative upper bound on the achievable rate with i.i.d. input symbols which is based on the one-step channel prediction error variance. The novelty lies in the fact that this bound is not restricted to peak power constrained input symbols like known bounds, e.g., in [1]. Therefore, the derived upper bound can also be used to evaluate the achievable rate with i.i.d. proper Gaussian input symbols. In addition, we compare the derived bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols with bounds on the peak power constrained capacity given in [1], [2], and [3]. Finally, we compare the achievable rate with i.i.d. zero-mean proper Gaussian input symbols with the achievable rate using synchronized detection in combination with a solely pilot based channel estimation.

## Index Terms

Channel capacity, fading channels, Gaussian distributions, information rates, noncoherent, Rayleigh, time-selective.

## I. INTRODUCTION

**I**N this paper, we consider a stationary Rayleigh flat-fading channel with temporal correlation. We assume that the channel state information is unknown to the transmitter and the receiver, while the receiver is aware of the channel law. The capacity of this scenario is particularly important, as it applies to many realistic mobile communication systems. In order to acquire channel state information, the temporal correlation of the channel can be exploited by the system, e.g., by inserting training sequences at transmit side. While these training sequences can be understood as a specific type of code [3], we are interested in the achievable rate on this channel irrespective of the use of training sequences.

The capacity of fading channels where the channel state information is unknown, i.e., sometimes referred to as noncoherent capacity, has received a lot of attention in the recent literature. E.g., [4] considers a block fading channel model, where the channel is assumed to be constant over a block of  $N$  symbols and changes independently from block to block. This model is non-stationary and, therefore, different

This work has been supported by the UMIC (Ultra High Speed Mobile Information and Communication) research centre. The material in this paper was presented in part at the International Symposium on Information Theory and its Applications (ISITA) 2008, Auckland, New Zealand, December 2008, and at the International Symposium on Information Theory and its Applications (ISITA) 2010, Taichung, Taiwan, October 2010.

M. Dörpinghaus was with the Institute for Integrated Signal Processing Systems, RWTH Aachen University, 52056 Aachen, Germany and is now with the Institute for Theoretical Information Technology, RWTH Aachen University, 52056 Aachen, Germany (e-mail: doerpinghaus@ti.rwth-aachen.de).

H. Meyr is with the Institute for Integrated Signal Processing Systems, RWTH Aachen University, 52056 Aachen, Germany (e-mail: meyr@iss.rwth-aachen.de).

from the one we consider in the present work. On the other hand, in [5] and [6] the achievable rate of time-continuous fading channels has been examined under the assumption of the use of training sequences for channel tracking and a coherent detection based on the acquired channel estimate. Furthermore, in [3] the asymptotic high SNR capacity of a stationary Gaussian fading channel has been investigated, whereas in [7] an approximate behavior of the capacity for different SNR regimes has been considered. In addition, in [1] bounds on the capacity for temporally correlated Rayleigh fading channels with a peak power constraint have been derived with specific emphasis on the low SNR regime. Furthermore, the case of frequency selective stationary fading channels has been discussed, e.g., in [8] and [9].

The main goal of the present work is the investigation of the achievable rate with independent identically distributed (i.i.d.) zero-mean proper Gaussian input symbols on noncoherent stationary discrete-time Rayleigh flat-fading channels. On the one hand, i.i.d. zero-mean proper Gaussian input symbols are capacity achieving in case the channel state is perfectly known at the receiver. Even though they are not capacity achieving for the given noncoherent scenario [10], the achievable rate with i.i.d. zero-mean proper Gaussian input symbols is highly interesting, as in many cases the capacity-achieving input distribution becomes peaky and, thus, impractical for real system design. In contrast, i.i.d. zero-mean proper Gaussian input distributions serve well to upper-bound the achievable rate with practical modulation and coding schemes, see also [11] and [12]. In [11] the achievable rate with i.i.d. Gaussian inputs has been studied for the block fading channel. Furthermore, in [13] bounds on the mutual information with Gaussian input distributions have been derived for a Gauss-Markov fading channel, whose PSD has an unbounded support. The results in [13] indicate that at moderate SNR and/or slow fading, Gaussian inputs still work well. In contrast to these publications, in the present work we study the achievable rate with i.i.d. zero-mean proper Gaussian input symbols for the case of a discrete-time stationary Rayleigh flat-fading channel, where the power spectral density (PSD) of the channel fading process is characterized by a compact support with a normalized maximum Doppler frequency  $f_d < 0.5$ , i.e., *nonregular* fading [14], as it, e.g., corresponds to the widely used Jakes' model [15].

#### A. Contributions

Within the present work, we consider a discrete-time stationary Rayleigh flat-fading channel with a compactly supported PSD. The channel fading process is assumed to be nonregular. Furthermore, it is assumed that the channel state information is unknown to the transmitter and the receiver, while the receiver is aware of the channel law. In this context we obtain the following.

In Section III, we give an upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols for the special case of a rectangular PSD, depending on the SNR and the spread of the PSD. Especially, the therefor used lower bound on the conditional output entropy rate  $h'(y|x)$  for a rectangular PSD is, to the best of our knowledge, new. The particularity of the given lower bound on  $h'(y|x)$  lies in the fact that its derivation is not based on a peak power constraint, enabling its evaluation for Gaussian input symbols. The assumption of a rectangular PSD is usually made in typical communication system design. For comparison, we give a lower bound on the achievable rate, holding for an arbitrary PSD with compact support, which is already known from [16]. With this lower and this upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs, we have found a set of bounds which is tight in the sense that its difference is bounded by  $(1 + 2f_d)\gamma$  [nat/channel use] for all SNRs, where  $\gamma \approx 0.57721$  is the Euler constant and  $f_d$  is the normalized maximum Doppler frequency. Furthermore, in Section III-G, we discuss the relation of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols to bounds on the peak power constrained capacity given in [2] and [1].

We show, in Section III-E, that the asymptotic high SNR slope (*pre-log*) of the achievable rate with i.i.d. zero-mean proper Gaussian input symbols is given by  $1 - 2f_d$ . This exactly corresponds to the high SNR behavior of the peak power constrained capacity as discussed in [3]. Furthermore, in Section III-F, we compare the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols to the high SNR asymptotes on the peak power constrained capacity given in [3].

Additionally, in Section IV, we derive an alternative upper bound on the achievable rate with i.i.d. input symbols. This upper bound, whose derivation relies on the assumption on i.i.d. input symbols, is based on the one-step channel prediction error variance, and, thus, is linked to a physical interpretation. Compared to this, the bounds given in Section III are based on a purely mathematical derivation. There exist already bounds on the capacity with peak power constrained input distributions, which are based on the one-step channel prediction error variance, see, e.g., [3]. However, for the derivation of the channel prediction based capacity bounds in [3], the peak power constraint has been required for technical reasons. Differently, our new upper bound based on the channel prediction error variance is not restricted to peak power constrained input symbols, enabling its evaluation for Gaussian inputs. However, due to the restriction to i.i.d. input symbols, we only get an upper bound on the achievable rate and not on the capacity. We evaluate this upper bound on the achievable rate with i.i.d. input symbols, on the one hand, for zero-mean proper Gaussian input symbols. In contrast to the upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols derived in Section III, which holds only for a rectangular PSD of the fading process, the upper bound based on the channel prediction error variance holds for an arbitrary PSD with compact support.

On the other hand, and this is out of the main focus of the present paper, we also evaluate the upper bound on the achievable rate with i.i.d. input symbols based on the channel prediction error variance for peak power constrained input symbols, and compare this new upper bound to capacity bounds given in [1].

Finally, we compare the achievable rate with i.i.d. zero-mean proper Gaussian input symbols to the achievable rate with synchronized detection and a solely pilot based channel estimation. Using synchronized detection with a solely pilot based channel estimation, the channel is estimated solely based on pilot symbols and then, in a second separate step, the channel estimate is used for coherent detection. Thus, this comparison shows, how far such systems stay below the achievable rate with Gaussian code books, using no pilot symbols. This comparison might give an indication of the possible gain of advanced receivers using a joint processing of pilot and data symbols, in comparison to the separate processing. Such an instance of a joint processing of pilots and data symbols is, e.g., the approach of iterative code-aided channel estimation and decoding [17], where the channel estimation is iteratively enhanced based on reliability information on the data symbols delivered by the decoder. The enhanced channel estimation, then in a further iteration allows for an enhanced decoding, and so on.

The rest of this paper is organized as follows. In Section II, we introduce the channel model including a discussion of its limitations. Section III presents the derivation of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, which are based on a purely mathematical derivation. This includes the discussion of their tightness, the evaluation of the high SNR behavior, the comparison to the high SNR asymptotes for the capacity given in [3], and the discussion of their relation to the bounds on the peak power constrained capacity given in [2] and [1]. Afterwards, in Section IV an upper bound on the achievable rate with i.i.d. input symbols based on the channel prediction error variance is derived and evaluated for proper Gaussian inputs and for peak power constrained inputs. Subsequently, in Section V the achievable rate with i.i.d. zero-mean proper Gaussian input symbols is compared to the achievable rate with synchronized detection and a solely pilot based channel estimation, before we give a conclusion in Section VI.

## II. SYSTEM MODEL

We consider an ergodic discrete-time jointly proper Gaussian [18] flat-fading channel, whose output at time  $k$  is given by

$$y_k = h_k \cdot x_k + n_k \quad (1)$$

where  $x_k \in \mathbb{C}$  is the complex-valued channel input,  $h_k \in \mathbb{C}$  represents the channel fading coefficient, and  $n_k \in \mathbb{C}$  is additive white Gaussian noise. The processes  $\{h_k\}$ ,  $\{x_k\}$ , and  $\{n_k\}$  are assumed to be mutually independent.

We assume that the noise  $\{n_k\}$  is a sequence of i.i.d. proper Gaussian random variables of zero-mean and variance  $\sigma_n^2$ . The stationary channel fading process  $\{h_k\}$  is zero-mean jointly proper Gaussian. In addition, the fading process is time-selective and characterized by its autocorrelation function

$$r_h(l) = \mathbb{E}[h_{k+l} \cdot h_k^*]. \quad (2)$$

Its variance is given by  $r_h(0) = \sigma_h^2$ .

The normalized PSD of the channel fading process is defined by

$$S_h(f) = \sum_{l=-\infty}^{\infty} r_h(l) e^{-j2\pi lf}, \quad |f| < 0.5 \quad (3)$$

where we assume that the PSD exists and use the definition  $j = \sqrt{-1}$ . Here, the frequency  $f$  is normalized with respect to the symbol duration  $T_{\text{Sym}}$ . In the following, we use this normalized PSD and, thus, refer to it as PSD for simplification. For a jointly proper Gaussian process, the existence of the PSD implies ergodicity [19]. As the channel fading process  $\{h_k\}$  is assumed to be stationary,  $S_h(f)$  is real-valued. Because of the limitation of the velocity of the transmitter, the receiver, and of objects in the environment, the spread of the PSD is limited, and we assume it to be compactly supported within the interval  $[-f_d, f_d]$ , with  $0 < f_d < 0.5$ , i.e.,  $S_h(f) = 0$  for  $f \notin [-f_d, f_d]$ . The parameter  $f_d$  corresponds to the normalized maximum Doppler shift and, thus, indicates the dynamics of the channel. To ensure ergodicity, we exclude the case  $f_d = 0$ . Following the definition given in [14], this fading channel is sometimes referred to as *nonregular*.

For technical reasons, in some of the proofs, i.e., for the calculation of the upper bound on the achievable data rate in Section III, we restrict to autocorrelation functions  $r_h(l)$  which are absolutely summable, i.e.,

$$\sum_{l=-\infty}^{\infty} |r_h(l)| < \infty \quad (4)$$

instead of the more general class of square summable autocorrelation functions, i.e.,

$$\sum_{l=-\infty}^{\infty} |r_h(l)|^2 < \infty. \quad (5)$$

The assumption of absolutely summable autocorrelation functions is not a severe restriction. E.g., the important rectangular PSD, see below in (7), can be arbitrarily closely approximated by a PSD with the shape corresponding to the transfer function of a raised cosine filter, whose autocorrelation function is absolutely summable, see Appendix A. Therefore, in the rest of this work, we often evaluate the derived bounds on the achievable rate for a rectangular PSD of the channel fading process, although some of the derivations are based on the assumption of an absolutely summable autocorrelation function.

A common model concerning the temporal correlation of the channel fading process  $S_h(f)$  is the Jakes' model [15], for which the corresponding PSD of the discrete-time fading process  $S_h(f)$  in (3) is given by

$$S_h(f)|_{\text{Jakes}} = \begin{cases} \frac{\sigma_h^2}{\pi \sqrt{f_d^2 - f^2}} & \text{for } |f| < f_d \\ 0 & \text{for } f_d \leq |f| \leq 0.5 \end{cases}. \quad (6)$$

This PSD can be derived analytically for a dense scatterer environment with a vertical receive antenna and a constant azimuthal gain, a uniform distribution of signals arriving at all angles with phases being independently distributed over all angles, i.e., in the interval  $[0, 2\pi)$ , based on a sum of sinusoids [15].

Often the Jakes' PSD in (6) is approximated by the following rectangular PSD

$$S_h(f)|_{\text{Rect}} = \begin{cases} \frac{\sigma_h^2}{2f_d} & \text{for } |f| \leq f_d \\ 0 & \text{for } f_d < |f| \leq 0.5 \end{cases}. \quad (7)$$

For the derivation of the upper bound on the achievable rate in Section III we restrict to rectangular PSDs for mathematical tractability.

Typical fading channels, as they are observed in mobile communication environments, are characterized by relatively small normalized Doppler frequencies  $f_d$  in the regime of  $f_d \ll 0.1$ . Therefore, the restriction to channels with  $f_d < 0.5$ , i.e., *nonregular* fading, in the present work is reasonable. Although in practical scenarios the observed channel dynamics are very small, within this work, we consider the range of  $0 < f_d < 0.5$  to get a thorough understanding of the behavior of the bounds on the achievable rate.

#### A. Matrix-Vector Notation

We base the derivation of bounds on the achievable rate on the following matrix-vector notation of the system model:

$$\mathbf{y} = \mathbf{H} \cdot \mathbf{x} + \mathbf{n} = \mathbf{X} \cdot \mathbf{h} + \mathbf{n} \quad (8)$$

where the vectors  $\mathbf{x}$  is defined as

$$\mathbf{x} = [x_1, \dots, x_N]^T. \quad (9)$$

The vectors  $\mathbf{y}$  and  $\mathbf{n}$  are defined analogously. The matrix  $\mathbf{H}$  is diagonal and defined as  $\mathbf{H} = \text{diag}(\mathbf{h})$  with  $\mathbf{h} = [h_1, \dots, h_N]^T$ . Here the  $\text{diag}(\cdot)$  operator generates a diagonal matrix whose diagonal elements are given by the argument vector. The diagonal matrix  $\mathbf{X}$  is given by  $\mathbf{X} = \text{diag}(\mathbf{x})$ . The quantity  $N$  is the number of considered symbols. Later on, we investigate the case of  $N \rightarrow \infty$  to evaluate the achievable rate.

Using this vector notation, we express the temporal correlation of the fading process by the correlation matrix

$$\mathbf{R}_h = \mathbb{E}[\mathbf{h}\mathbf{h}^H] \quad (10)$$

which has a Hermitian Toeplitz structure.

Concerning the input distribution, unless otherwise stated, we make the assumption that the symbols  $x_k$  are i.i.d. zero-mean proper Gaussian distributed with an average power  $\sigma_x^2$ . Thus, the average SNR is given by<sup>1</sup>

$$\rho = \frac{\sigma_x^2 \sigma_h^2}{\sigma_n^2}. \quad (11)$$

#### B. Limitations of the Symbol Rate Discrete-Time Model

To discuss the limitations of the symbol rate discrete-time model given in (1), we start from the underlying appropriately bandlimited continuous-time model, where the channel output is given by

$$y(t) = h(t) \cdot s(t) + n(t) \quad (12)$$

with  $h(t)$  being the continuous-time channel fading process, i.e., the corresponding discrete-time process  $h_k$  is given by

$$h_k = h(kT_{\text{Sym}}) \quad (13)$$

<sup>1</sup>Remark: Only in case there is no peak power constraint, as, e.g., in the case of Gaussian input symbols, the average SNR  $\rho$  is in general equal to the actual average SNR. In contrast, in case of an additional peak power constraint the achievable rate is in general not maximized by using the maximum average transmit power  $\sigma_x^2$ . Thus, in this case  $\rho$  does not necessarily correspond to the actual average SNR and, therefore, we name it *nominal* average SNR when considering a peak power constraint. For further discussion see Section III-F and Section III-G1.

where  $T_{\text{Sym}}$  is the symbol duration. Analogously, the continuous-time and the discrete-time additive noise and channel output processes are related by

$$n_k = n(kT_{\text{Sym}}) \quad (14)$$

$$y_k = y(kT_{\text{Sym}}). \quad (15)$$

The continuous-time transmit process  $s(t)$  is given by

$$s(t) = \sum_{k=-\infty}^{\infty} x_k \cdot g(t - kT_{\text{Sym}}) \quad (16)$$

where  $g(t)$  is the transmit pulse. We assume the use of bandlimited transmit pulses, which, therefore, have an infinite impulse response. In typical systems, often root-raised cosine pulses are used such that in combination with the matched filter at the receiver intersymbol interference is minimized. Their normalized frequency response  $G(f)$  is given by

$$G(f) = \sqrt{G_{\text{RC}}(f)} \quad (17)$$

with  $G_{\text{RC}}(f)$  being the transfer function of the raised cosine filter

$$G_{\text{RC}}(f) = \begin{cases} T_{\text{Sym}} & \text{for } |f| \leq \frac{1-\beta_{\text{ro}}}{2} \\ \frac{T_{\text{Sym}}}{2} \left[ 1 + \cos \left( \frac{\pi}{\beta_{\text{ro}}} \left[ |f| - \frac{1-\beta_{\text{ro}}}{2} \right] \right) \right] & \text{for } \frac{1-\beta_{\text{ro}}}{2} < |f| \leq \frac{1+\beta_{\text{ro}}}{2} \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

where  $0 \leq \beta_{\text{ro}} \leq 1$  is the roll-off factor.

The continuous-time input/output relation in (12) has the following stochastic representation in frequency domain

$$S_y(f) = S_h(f) \star S_s(f) + S_n(f) \quad (19)$$

where  $\star$  denotes convolution and  $S_y(f)$ ,  $S_h(f)$ ,  $S_s(f)$ , and  $S_n(f)$  are the normalized power spectral densities of the continuous-time processes  $y(t)$ ,  $h(t)$ ,  $s(t)$ , and  $n(t)$ , e.g.,

$$S_s(f) = \int_{-\infty}^{\infty} \mathbb{E} [s(t + \tau) s^*(t)] e^{-j2\pi f \tau} d\tau \quad (20)$$

and correspondingly for the other PSDs. Here, we always assume normalization with  $1/T_{\text{Sym}}$ .

We are interested in the normalized bandwidth of the component  $S_h(f) \star S_s(f)$ , i.e., the component containing information on the transmitted sequence  $\{x_k\}$ . The normalized bandwidth of the transmit signal  $s(t)$  directly corresponds to the normalized bandwidth of the transmit pulse  $g(t)$ , which is at least 1 (corresponding to  $\beta_{\text{ro}} = 0$ ). The normalized bandwidth of the channel fading process is given by  $2f_d$ . Thus, the normalized bandwidth of the component  $S_h(f) \star S_s(f)$  is given by at least  $1 + 2f_d$ . To get a sufficient statistic, we would have to sample the channel output  $y(t)$  at least with a frequency of  $\frac{1+2f_d}{T_{\text{Sym}}}$  (for  $\beta_{\text{ro}} = 0$ ). As the discrete-time channel output process  $\{y_k\}$  is a sampled version of  $y(t)$  with the rate  $1/T_{\text{Sym}}$ , the discrete-time observation process  $\{y_k\}$  is not a sufficient statistic of  $y(t)$ . This shows the limitations of the symbol rate discrete-time system model in (1). As in typical systems the normalized maximum Doppler frequency  $f_d$  is small in comparison to the symbol rate  $1/T_{\text{Sym}}$ , the amount of discarded information is negligible. Besides this, in typical systems channel estimation is also performed at symbol rate and, therefore, also exhibit the loss due to the lack of a sufficient statistic. In addition, the majority of the current literature on the study of the capacity of stationary Rayleigh fading channels, e.g., [3] or [1], is based on symbol rate discrete-time input-output relations and therefore do not ask the question about a sufficient statistic. Nevertheless, these considerations should be kept in mind in the following, especially, as we examine the derived bounds not only for very small values of  $f_d$ .

### III. BOUNDS ON THE ACHIEVABLE RATE

The information theoretic capacity is defined by

$$C = \lim_{N \rightarrow \infty} \sup_{\mathcal{P}} \frac{1}{N} \mathcal{I}(\mathbf{y}; \mathbf{x}) \quad [\text{nat/cu}] \quad (21)$$

where  $\mathcal{I}(\mathbf{y}; \mathbf{x})$  is the mutual information between  $\mathbf{x}$  and  $\mathbf{y}$ , where *cu* is an abbreviation for *channel use*, and where the supremum is taken over the set  $\mathcal{P}$  containing all input distributions with an average power constrained to  $\sigma_x^2$ , i.e.,

$$\mathcal{P} = \left\{ p(\mathbf{x}) \left| \mathbf{x} \in \mathbb{C}^N, \frac{1}{N} \mathbb{E} [\mathbf{x}^H \mathbf{x}] \leq \sigma_x^2 \right. \right\}. \quad (22)$$

As the PSD of the fading process in (3) is assumed to exist, and as the channel fading process is jointly proper Gaussian, the channel fading process is ergodic. Therefore, the information theoretic capacity given in (21) and the operational capacity coincide [19], i.e., for each rate  $R < C$  there exist a code for which the probability of an erroneously decoded codeword approaches zero in the limit of an infinite codeword length.

As it has already been discussed, the main focus of the present paper is not the discussion of the capacity but of the achievable rate with i.i.d. zero-mean proper Gaussian input symbols. As this kind of input distribution is in general not capacity achieving, we use the term *achievable rate*  $R$ , which then directly corresponds to the mutual information rate  $\mathcal{I}'(\mathbf{y}; \mathbf{x})$ , i.e.,

$$R = \mathcal{I}'(\mathbf{y}; \mathbf{x}) = \lim_{N \rightarrow \infty} \frac{1}{N} \mathcal{I}(\mathbf{y}; \mathbf{x}) \quad (23)$$

where, as described in Section II, the elements of  $\mathbf{x}$  are i.i.d. zero-mean proper Gaussian distributed.

#### A. The Mutual Information Rate $\mathcal{I}'(\mathbf{y}; \mathbf{x})$

In general, by means of the chain rule, the mutual information rate in (23) can be expanded as [20]

$$\mathcal{I}'(\mathbf{y}; \mathbf{x}) = \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) - \mathcal{I}'(\mathbf{x}; \mathbf{h}|\mathbf{y}) \quad (24)$$

where  $\mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$  is the mutual information rate in case the channel is known at the receiver, i.e., the mutual information rate of the coherent channel, and  $\mathcal{I}'(\mathbf{x}; \mathbf{h}|\mathbf{y})$  is the penalty due to the channel uncertainty. It is interesting to note that the penalty term can be further separated as follows:

$$\begin{aligned} \mathcal{I}'(\mathbf{x}; \mathbf{h}|\mathbf{y}) &\stackrel{(a)}{=} \mathcal{I}'(\mathbf{y}, \mathbf{x}; \mathbf{h}) - \mathcal{I}'(\mathbf{y}; \mathbf{h}) \\ &\stackrel{(b)}{=} \mathcal{I}'(\mathbf{y}; \mathbf{h}|\mathbf{x}) + \mathcal{I}'(\mathbf{h}; \mathbf{x}) - \mathcal{I}'(\mathbf{y}; \mathbf{h}) \\ &\stackrel{(c)}{=} \mathcal{I}'(\mathbf{y}; \mathbf{h}|\mathbf{x}) - \mathcal{I}'(\mathbf{y}; \mathbf{h}) \end{aligned} \quad (25)$$

where for (a) and (b) we use the chain rule for mutual information and for (c) we exploit the fact that the mutual information between the channel fading process described by  $\mathbf{h}$  and the input sequence  $\mathbf{x}$  is zero due to the independency of  $\mathbf{h}$  and  $\mathbf{x}$  and, thus,

$$\mathcal{I}'(\mathbf{h}; \mathbf{x}) = 0. \quad (26)$$

Obviously, with (25) the penalty term corresponds to the difference between the knowledge on the channel  $\mathbf{h}$  that can be obtained from the observation  $\mathbf{y}$  while knowing the transmit sequence  $\mathbf{x}$  in comparison to not knowing it.

However, the derivation of the bounds on the mutual information rate  $\mathcal{I}'(\mathbf{y}; \mathbf{x})$  in the present work is based on the following straightforward separation of  $\mathcal{I}'(\mathbf{y}; \mathbf{x})$  into the differential entropy rates

$$\mathcal{I}'(\mathbf{y}; \mathbf{x}) = h'(\mathbf{y}) - h'(\mathbf{y}|\mathbf{x}). \quad (27)$$

where  $h'(\cdot)$  denotes the differential entropy rate

$$h'(\cdot) = \lim_{N \rightarrow \infty} \frac{1}{N} h(\cdot). \quad (28)$$

In Section III-B, we give a lower and an upper bound on the channel output entropy rate  $h'(\mathbf{y})$ , which are independent of the PSD of the channel fading process  $S_h(f)$ . In Section III-C, we derive an upper bound and the lower bound on  $h'(\mathbf{y}|\mathbf{x})$ . The upper bound, which is already known from [16], holds for an arbitrary PSD of the channel fading process with compact support. For the lower bound on  $h'(\mathbf{y}|\mathbf{x})$  we find a closed form expression only for the special case of a rectangular PSD.

### B. The Received Signal Entropy Rate $h'(\mathbf{y})$

1) *Lower Bound on  $h'(\mathbf{y})$* : The mutual information with perfect channel state information at the receiver can be upper-bounded by

$$\begin{aligned} \mathcal{I}(\mathbf{y}; \mathbf{x}|\mathbf{h}) &= h(\mathbf{y}|\mathbf{h}) - h(\mathbf{y}|\mathbf{h}, \mathbf{x}) \\ &\leq h(\mathbf{y}) - h(\mathbf{y}|\mathbf{h}, \mathbf{x}). \end{aligned} \quad (29)$$

Here, we make use of the fact that conditioning reduces entropy. Thus, we can lower-bound the entropy rate  $h'(\mathbf{y})$  by

$$h'(\mathbf{y}) \geq \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) + h'(\mathbf{y}|\mathbf{h}, \mathbf{x}). \quad (30)$$

The mutual information rate in case the channel is known at the receiver, i.e., the first term on the RHS of (30), is given by

$$\begin{aligned} \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) &= \frac{1}{N} \mathbb{E}_{\mathbf{h}} \left[ \mathbb{E}_{\mathbf{y}, \mathbf{x}} \left[ \log \left( \frac{p(\mathbf{y}|\mathbf{h}, \mathbf{x})}{p(\mathbf{y}|\mathbf{h})} \right) \middle| \mathbf{h} \right] \right] \\ &\stackrel{(a)}{=} \mathbb{E}_{h_k} \left[ \mathbb{E}_{y_k, x_k} \left[ \log \left( \frac{p(y_k|h_k, x_k)}{p(y_k|h_k)} \right) \middle| h_k \right] \right] \\ &= \mathcal{I}(y; x|h) \\ &\stackrel{(b)}{=} \mathbb{E}_{h_k} \left[ \log \left( 1 + \rho \frac{|h_k|^2}{\sigma_h^2} \right) \right] \\ &= \int_{z=0}^{\infty} \log(1 + \rho z) e^{-z} dz \end{aligned} \quad (31)$$

where (a) is based on the fact that due to conditioning on the channel fading vector  $\mathbf{h}$  the channel uses become independent and we furthermore assume i.i.d. input symbols.<sup>2</sup> Therefore, we can drop the time index for ease of notation. Finally, (b) holds for i.i.d. zero-mean proper Gaussian inputs which are capacity achieving in the coherent case. Thus, the RHS of (31) is the capacity of the coherent channel. Obviously, the coherent capacity is independent of the temporal correlation of the channel, see, e.g., [6].

The second term on the RHS of (30) originates from AWGN and, thus, can be calculated as

$$h'(\mathbf{y}|\mathbf{h}, \mathbf{x}) = \log(\pi e \sigma_n^2). \quad (32)$$

Hence, a lower bound on the entropy rate  $h'(\mathbf{y})$  is given by

$$h'(\mathbf{y}) \geq h'_L(\mathbf{y}) = \int_{z=0}^{\infty} \log[\pi e (\sigma_n^2 + \sigma_x^2 \sigma_h^2 z)] e^{-z} dz. \quad (33)$$

<sup>2</sup>All logarithms in this paper are to the base  $e$  and, unless stated otherwise, all rates are in *nat*.



2) *Upper Bound on  $h'(\mathbf{y})$* : In this section, we give an upper bound on the entropy rate  $h'(\mathbf{y})$ . First, we make use of the fact that the entropy  $h(\mathbf{y})$  of a zero-mean complex random vector  $\mathbf{y}$  of dimension  $N$  with nonsingular correlation matrix  $\mathbf{R}_y = \mathbb{E}[\mathbf{y}\mathbf{y}^H]$  is upper-bounded by [18]

$$\begin{aligned} h(\mathbf{y}) &\leq \log [(\pi e)^N \det(\mathbf{R}_y)] \\ &\stackrel{(a)}{=} N \log [\pi e (\sigma_x^2 \sigma_h^2 + \sigma_n^2)] . \end{aligned} \quad (34)$$

Here (a) follows from the fact that  $\mathbf{R}_y$  is diagonal and given by

$$\mathbf{R}_y = (\sigma_x^2 \sigma_h^2 + \sigma_n^2) \mathbf{I}_N. \quad (35)$$

due to the assumption on i.i.d. input symbols. Nevertheless, the upper bound on  $h(\mathbf{y})$  in (34) also holds without the assumption on independent input symbols, which can be easily verified by using Hadamard's inequality instead of the equality in (a).

Hence, with (34) the upper bound  $h'_U(\mathbf{y})$  on the entropy rate  $h'(\mathbf{y})$  is given by

$$h'(\mathbf{y}) \leq h'_U(\mathbf{y}) = \log (\pi e (\sigma_x^2 \sigma_h^2 + \sigma_n^2)) . \quad (36)$$

In Appendix B, we give another upper bound on  $h'(\mathbf{y})$  for the case of zero-mean proper Gaussian inputs based on numerical integration to calculate  $h(y_k)$ , i.e., the output entropy at an individual time instant, see also [21]. As this bound can only be evaluated numerically using Hermite polynomials and Simpson's rule or by Monte Carlo integration, we do not further consider it here.

### C. The Entropy Rate $h'(\mathbf{y}|\mathbf{x})$

In this section, we give an upper bound and a lower bound on the conditional channel output entropy rate  $h'(\mathbf{y}|\mathbf{x})$ . The probability density of  $\mathbf{y}$  conditioned on  $\mathbf{x}$  is zero-mean proper Gaussian. Therefore, its entropy is

$$h(\mathbf{y}|\mathbf{x}) = \mathbb{E}_{\mathbf{x}} [\log ((\pi e)^N \det(\mathbf{R}_{y|x}))] \quad (37)$$

where the covariance matrix  $\mathbf{R}_{y|x}$  is given by

$$\begin{aligned} \mathbf{R}_{y|x} &= \mathbb{E}_{\mathbf{h}, \mathbf{n}} [\mathbf{y}\mathbf{y}^H | \mathbf{x}] \\ &= \mathbb{E}_{\mathbf{h}} [\mathbf{X}\mathbf{h}\mathbf{h}^H \mathbf{X}^H | \mathbf{x}] + \sigma_n^2 \mathbf{I}_N \\ &= \mathbf{X}\mathbf{R}_h \mathbf{X}^H + \sigma_n^2 \mathbf{I}_N. \end{aligned} \quad (38)$$

As the channel correlation matrix  $\mathbf{R}_h$  is Hermitian and, thus, normal, the spectral decomposition theorem applies, i.e.,

$$\mathbf{R}_h = \mathbf{U} \mathbf{\Lambda}_h \mathbf{U}^H \quad (39)$$

where the diagonal matrix  $\mathbf{\Lambda}_h = \text{diag}(\lambda_1, \dots, \lambda_N)$  contains the eigenvalues  $\lambda_i$  of  $\mathbf{R}_h$  and the matrix  $\mathbf{U}$  is unitary.

1) *Upper Bound on  $h'(\mathbf{y}|\mathbf{x})$* : The following upper-bounding of  $h(\mathbf{y}|\mathbf{x})$  is already known from [16]. Making use of (39), Jensen's inequality and the concavity of the log function, we can upper-bound  $h(\mathbf{y}|\mathbf{x})$  in (37) as follows:

$$h(\mathbf{y}|\mathbf{x}) = \mathbb{E}_{\mathbf{x}} \left[ \log \det \left( \frac{1}{\sigma_n^2} \mathbf{X} \mathbf{U} \mathbf{\Lambda}_h \mathbf{U}^H \mathbf{X}^H + \mathbf{I}_N \right) \right] + N \log(\pi e \sigma_n^2) \quad (40)$$

$$\stackrel{(a)}{=} \mathbb{E}_{\mathbf{x}} \left[ \log \det \left( \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{U} \mathbf{\Lambda}_h \mathbf{U}^H + \mathbf{I}_N \right) \right] + N \log(\pi e \sigma_n^2) \quad (41)$$

$$\begin{aligned} &\stackrel{(b)}{\leq} \log \det \left( \frac{\sigma_x^2}{\sigma_n^2} \mathbf{U} \mathbf{\Lambda}_h \mathbf{U}^H + \mathbf{I}_N \right) + N \log(\pi e \sigma_n^2) \\ &= \log \det \left( \frac{\sigma_x^2}{\sigma_n^2} \mathbf{\Lambda}_h + \mathbf{I}_N \right) + N \log(\pi e \sigma_n^2) \\ &= \sum_{i=1}^N \log \left( \frac{\sigma_x^2}{\sigma_n^2} \lambda_i + 1 \right) + N \log(\pi e \sigma_n^2). \end{aligned} \quad (42)$$

For (a) the following relation is used

$$\det(\mathbf{A}\mathbf{B} + \mathbf{I}) = \det(\mathbf{B}\mathbf{A} + \mathbf{I}) \quad (43)$$

which holds as  $\mathbf{A}\mathbf{B}$  has the same eigenvalues as  $\mathbf{B}\mathbf{A}$  for  $\mathbf{A}$  and  $\mathbf{B}$  being square matrices [22, Theorem 1.3.20]. For (b) we have used the fact that  $\log \det(\cdot)$  is concave on the set of positive definite matrices<sup>3</sup>.

To calculate the bound on the entropy rate  $h'(\mathbf{y}|\mathbf{x})$ , we consider the case  $N \rightarrow \infty$ , i.e., the dimension of the matrix  $\mathbf{\Lambda}_h$  grows without bound. As  $\mathbf{R}_h$  is Hermitian Toeplitz, we can evaluate (42) using Szegő's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [23], [24]. Consequently,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \log \left( \frac{\sigma_x^2}{\sigma_n^2} \lambda_i + 1 \right) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left( S_h(f) \frac{\sigma_x^2}{\sigma_n^2} + 1 \right) df. \quad (44)$$

Notice that due to the assumption that the PSD exists, the autocorrelation function of the channel fading process is square summable, see (5), and, thus, Szegő's theorem can be applied.

Hence, we get the following upper bound:

$$\begin{aligned} h'(\mathbf{y}|\mathbf{x}) &\leq h'_U(\mathbf{y}|\mathbf{x}) \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left( S_h(f) \frac{\sigma_x^2}{\sigma_n^2} + 1 \right) df + \log(\pi e \sigma_n^2). \end{aligned} \quad (45)$$

At this point, it is interesting to note that for constant modulus (CM) input symbols the differential entropy rate  $h'(\mathbf{y}|\mathbf{x})$  is equal to the upper bound  $h'_U(\mathbf{y}|\mathbf{x})$ , i.e.,

$$h'(\mathbf{y}|\mathbf{x})|_{\text{CM}} = h'_U(\mathbf{y}|\mathbf{x}) \quad (46)$$

as in this case (41) simplifies as

$$\mathbf{X}^H \mathbf{X}|_{\text{CM}} = \sigma_x^2 \mathbf{I} \quad (47)$$

and, thus, (b) succeeding (41) holds with equality.

<sup>3</sup>For the special case of independent transmit symbols, (b) can also be shown in two steps by using Jensen's inequality and in a second step expressing the determinant by a Laplacian expansion by minors to calculate the expectation, i.e.,

$$\begin{aligned} \mathbb{E}_{\mathbf{x}} \left[ \log \det \left( \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{U} \mathbf{\Lambda}_h \mathbf{U}^H + \mathbf{I}_N \right) \right] &\leq \log \mathbb{E}_{\mathbf{x}} \left[ \det \left( \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{U} \mathbf{\Lambda}_h \mathbf{U}^H + \mathbf{I}_N \right) \right] \\ &= \log \det \left( \frac{\sigma_x^2}{\sigma_n^2} \mathbf{U} \mathbf{\Lambda}_h \mathbf{U}^H + \mathbf{I}_N \right). \end{aligned}$$

2) *Lower Bound on  $h'(\mathbf{y}|\mathbf{x})$  for a Rectangular PSD:* In this section, we give a new lower bound on the entropy rate  $h'(\mathbf{y}|\mathbf{x})$  for the special case of a rectangular PSD, which is a common approximation of the actual PSD in typical system design.

For the derivation of this lower bound, we derive a circulant matrix which is asymptotically equivalent to the Toeplitz matrix  $\mathbf{R}_h$ . Hereby, we follow a specific approach as shown in [23], where the circulant matrix is constructed by sampling the PSD of the channel fading process. For the discussion of asymptotical equivalency, we write  $\mathbf{R}_h^{(N)}$  instead of  $\mathbf{R}_h$ , where the superscript  $(N)$  denotes the size of the square matrix  $\mathbf{R}_h$ .

Let the first column of the circulant matrix  $\mathbf{C}_h^{(N)}$  be given by

$$\begin{pmatrix} c_0^{(N)} & c_1^{(N)} & \dots & c_{N-1}^{(N)} \end{pmatrix}^T \quad (48)$$

where again the superscript  $(N)$  denotes the size of the square matrix  $\mathbf{C}_h^{(N)}$ . The elements  $c_k^{(N)}$  are given by

$$c_k^{(N)} = \frac{1}{N} \sum_{l=0}^{N-1} \tilde{S}_h\left(\frac{l}{N}\right) e^{j2\pi k \frac{l}{N}} \quad (49)$$

where  $\tilde{S}_h(f)$  is the periodic continuation of  $S_h(f)$  given in (3), i.e.,

$$\tilde{S}_h(f) = \sum_{k=-\infty}^{\infty} \delta(f - k) \star S_h(f) \quad (50)$$

and  $S_h(f)$  being zero outside the interval  $|f| \leq 0.5$  for which it is defined in (3).

As we assume that the autocorrelation function of the channel fading process is absolutely summable, the PSD of the channel fading process  $\tilde{S}_h(f)$  is Riemann integrable, and it holds that

$$\begin{aligned} \lim_{N \rightarrow \infty} c_k^{(N)} &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{l=0}^{N-1} \tilde{S}_h\left(\frac{l}{N}\right) e^{j2\pi k \frac{l}{N}} \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} S_h(f) e^{j2\pi k f} df = r_h(k) \end{aligned} \quad (51)$$

with  $r_h(k)$  given by (2).

As the eigenvectors of a circulant matrix are given by a discrete Fourier transform (DFT), the eigenvalues  $\tilde{\lambda}_k^{(N)}$  with  $k = 1, \dots, N$  of the circulant matrix  $\mathbf{C}_h^{(N)}$  are given by

$$\begin{aligned} \tilde{\lambda}_k^{(N)} &= \sum_{l=0}^{N-1} c_l^{(N)} e^{-j2\pi \frac{(k-1)l}{N}} \\ &= \sum_{l=0}^{N-1} \left( \frac{1}{N} \sum_{m=0}^{N-1} \tilde{S}_h\left(\frac{m}{N}\right) e^{j2\pi \frac{ml}{N}} \right) e^{-j2\pi \frac{l(k-1)}{N}} \\ &= \sum_{m=0}^{N-1} \tilde{S}_h\left(\frac{m}{N}\right) \left\{ \frac{1}{N} \sum_{l=0}^{N-1} e^{j2\pi \frac{l(m-(k-1))}{N}} \right\} \\ &= \tilde{S}_h\left(\frac{k-1}{N}\right). \end{aligned} \quad (52)$$

Consequently, the spectral decomposition of the circulant matrix  $\mathbf{C}_h^{(N)}$  is given by

$$\mathbf{C}_h^{(N)} = \mathbf{F}^{(N)} \tilde{\mathbf{\Lambda}}_h^{(N)} (\mathbf{F}^{(N)})^H \quad (53)$$

where the matrix  $\mathbf{F}^{(N)}$  is a unitary DFT matrix, i.e., its elements are given by

$$[\mathbf{F}^{(N)}]_{k,l} = \frac{1}{\sqrt{N}} e^{j2\pi \frac{(k-1)(l-1)}{N}}. \quad (54)$$

Furthermore, the matrix  $\tilde{\Lambda}_h^{(N)}$  is diagonal with the elements  $\tilde{\lambda}_k^{(N)}$  given in (52).

In [23, Lemma 4.6] it is shown that the circulant matrix  $\mathbf{C}_h^{(N)}$  and the Toeplitz matrix  $\mathbf{R}_h^{(N)}$  are asymptotically equivalent if the autocorrelation function  $r_h(l)$  is absolutely summable. In the context of proving this lemma it is shown that the weak norm of the difference of  $\mathbf{R}_h^{(N)}$  and  $\mathbf{C}_h^{(N)}$  converges to zero as  $N \rightarrow \infty$ , i.e.,

$$\lim_{N \rightarrow \infty} \|\mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)}\| = 0 \quad (55)$$

where the weak norm of a matrix  $\mathbf{B}$  is defined as

$$\|\mathbf{B}\| = \left( \frac{1}{N} \text{Tr} [\mathbf{B}^H \mathbf{B}] \right)^{\frac{1}{2}}. \quad (56)$$

The convergence of the weak norm of the difference  $\mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)}$  towards zero is required later on.

By the construction of the circulant matrix  $\mathbf{C}_h^{(N)}$ , the eigenvalues  $\tilde{\lambda}_k$  of  $\mathbf{C}_h^{(N)}$  are given by (52), i.e.,

$$\tilde{\lambda}_k^{(N)} = \begin{cases} S_h\left(f = \frac{k-1}{N}\right) & \text{for } 1 \leq k \leq \lceil \frac{N}{2} \rceil \\ S_h\left(f = \frac{k-1}{N} - 1\right) & \text{for } \lceil \frac{N}{2} \rceil < k \leq N \end{cases}. \quad (57)$$

Thus, if the PSD of the channel fading process  $S_h(f)$  is rectangular, the eigenvalues of the circulant matrix  $\mathbf{C}_h^{(N)}$  are given by

$$\tilde{\lambda}_k^{(N)} = \begin{cases} \frac{\sigma_h^2}{2f_d} & \text{for } 1 \leq k \leq f_d N + 1 \\ \quad \vee \quad (1 - f_d)N + 1 \leq k \leq N \\ 0 & \text{otherwise} \end{cases}. \quad (58)$$

This means that the eigenvalues of  $\mathbf{R}_h^{(N)}$  corresponding to frequencies  $|f| > f_d$  become zero for  $N \rightarrow \infty$ .

Now, we apply the asymptotic equivalence of  $\mathbf{R}_h^{(N)}$  and  $\mathbf{C}_h^{(N)}$  to lower-bound the entropy rate  $h'(\mathbf{y}|\mathbf{x})$  given by

$$h'(\mathbf{y}|\mathbf{x}) = \lim_{N \rightarrow \infty} \frac{1}{N} h(\mathbf{y}|\mathbf{x}) \quad (59)$$

with  $h(\mathbf{y}|\mathbf{x})$  given in (41). Thus, we have to show that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{\mathbf{x}} \left[ \log \det \left( \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{R}_h^{(N)} + \mathbf{I}_N \right) \right] = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{\mathbf{x}} \left[ \log \det \left( \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{C}_h^{(N)} + \mathbf{I}_N \right) \right]. \quad (60)$$

To prove (60), we have to show that the matrices

$$\mathbf{K}_1^{(N)} = \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{R}_h^{(N)} + \mathbf{I}_N \quad (61)$$

$$\mathbf{K}_2^{(N)} = \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{C}_h^{(N)} + \mathbf{I}_N \quad (62)$$

are asymptotically equivalent [23, Theorem 2.4]. This means that we have to show that both matrices are bounded in the strong norm, and that the weak norm of their difference converges to zero for  $N \rightarrow \infty$  [23, Section 2.3].

Concerning the condition with respect to the strong norm we have to show that

$$\left\| \mathbf{K}_1^{(N)} \right\| < \infty \quad (63)$$

$$\left\| \mathbf{K}_2^{(N)} \right\| < \infty \quad (64)$$

with the strong norm of the matrix  $\mathbf{B}$  defined by

$$\|\mathbf{B}\|^2 = \max_k \gamma_k \quad (65)$$

where  $\gamma_k$  are the eigenvalues of the Hermitian nonnegative definite matrix  $\mathbf{B}\mathbf{B}^H$ . The diagonal matrix  $\mathbf{X}^H\mathbf{X}$  contains the transmit powers of the individual transmit symbols. In the case of Gaussian input distributions, for a given  $\epsilon > 0$ , there exists a finite value  $M(\epsilon)$  such that the transmit power is smaller than  $M(\epsilon)$  with probability  $1 - \epsilon$ . In addition, the strong norms of  $\mathbf{R}_h^{(N)}$  and  $\mathbf{C}_h^{(N)}$  are bounded, too. Concerning the boundedness of the eigenvalues of the Hermitian Toeplitz matrix  $\mathbf{R}_h^{(N)}$  see [23, Lemma 4.1]. Thus, the strong norms of  $\mathbf{K}_1^{(N)}$  and  $\mathbf{K}_2^{(N)}$  are asymptotically almost surely bounded.

Furthermore, for the weak norm of the difference  $\mathbf{K}_1^{(N)} - \mathbf{K}_2^{(N)}$  we get for  $N \rightarrow \infty$

$$\begin{aligned} \left| \mathbf{K}_1^{(N)} - \mathbf{K}_2^{(N)} \right| &= \left| \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \left( \mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right) \right| \\ &\stackrel{(a)}{\leq} \frac{1}{\sigma_n^2} \left\| \mathbf{X}^H \mathbf{X} \right\| \left| \mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right| \end{aligned} \quad (66)$$

where for (a) we have used [23, Lemma 2.3].

Based on the above argumentation that  $\left\| \mathbf{X}^H \mathbf{X} \right\|$  is asymptotically almost surely bounded, we get for  $N \rightarrow \infty$

$$\begin{aligned} \lim_{N \rightarrow \infty} \left| \mathbf{K}_1^{(N)} - \mathbf{K}_2^{(N)} \right| &\leq \lim_{N \rightarrow \infty} \frac{1}{\sigma_n^2} \left\| \mathbf{X}^H \mathbf{X} \right\| \left| \mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right| \\ &= 0 \end{aligned} \quad (67)$$

due to (55). Thus, we have proved that (60) holds and we can express the entropy rate  $h'(\mathbf{y}|\mathbf{x})$  by

$$\begin{aligned} h'(\mathbf{y}|\mathbf{x}) &= \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{\mathbf{x}} \left[ \log \det \left( \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{C}_h^{(N)} + \mathbf{I}_N \right) \right] + \log(\pi e \sigma_n^2) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{\mathbf{x}} \left[ \log \det \left( \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{F} \tilde{\Lambda}_h \mathbf{F}^H + \mathbf{I}_N \right) \right] + \log(\pi e \sigma_n^2). \end{aligned} \quad (68)$$

Here,  $\mathbf{F} \tilde{\Lambda}_h \mathbf{F}^H$  is the spectral decomposition of the circulant matrix  $\mathbf{C}_h^{(N)}$ , see (53) (from here on we again omit the superscript  $(N)$  for ease of notation). Thus,  $\tilde{\Lambda}_h$  is a diagonal matrix containing the eigenvalues  $\tilde{\lambda}_k$  as given in (58) and the matrix  $\mathbf{F}$  is a unitary matrix with the eigenvectors of  $\mathbf{C}_h^{(N)}$  on its columns.

To calculate a lower bound on  $h'(\mathbf{y}|\mathbf{x})$ , we transform the term with the expectation operation at the RHS of (68) as follows:

$$\begin{aligned} \mathbb{E}_{\mathbf{x}} \left[ \log \det \left( \frac{1}{\sigma_n^2} \mathbf{X}^H \mathbf{X} \mathbf{F} \tilde{\Lambda}_h \mathbf{F}^H + \mathbf{I}_N \right) \right] &\stackrel{(a)}{=} \mathbb{E}_{\mathbf{x}} \left[ \log \det \left( \frac{1}{\sigma_n^2} \tilde{\Lambda}_h \mathbf{F}^H \mathbf{X}^H \mathbf{X} \mathbf{F} + \mathbf{I}_N \right) \right] \\ &\stackrel{(b)}{=} \mathbb{E}_{\mathbf{x}} \left[ \log \det \left( \frac{\sigma_h^2}{2f_d \sigma_n^2} \tilde{\mathbf{F}}^H \mathbf{X}^H \mathbf{X} \tilde{\mathbf{F}} + \mathbf{I}_{2[f_d N] + 1} \right) \right] \end{aligned} \quad (69)$$

where for (a) we have used (43). For (b) the eigenvalue distribution in (58) is used, and the matrix  $\tilde{\mathbf{F}}$  is given by

$$\tilde{\mathbf{F}} = [\mathbf{f}_1, \dots, \mathbf{f}_{[f_d N] + 1}, \mathbf{f}_{[(1-f_d)N] + 1}, \dots, \mathbf{f}_N] \in \mathbb{C}^{N \times (2[f_d N] + 1)} \quad (70)$$

where the  $\mathbf{f}_i$  are the orthonormal columns of the unitary matrix  $\mathbf{F}$ . I.e.,  $\tilde{\mathbf{F}}$  contains the eigenvectors corresponding to the non-zero eigenvalues of  $\mathbf{C}_h^{(N)}$ . Now, we apply the following inequality given in [25, Lemma 1].

*Lemma 1:* Let  $\mathbf{A} \in \mathbb{C}^{m \times n}$  with orthonormal rows and  $m \leq n$ . Then

$$\log \det (\mathbf{A} \text{diag}(p_1, \dots, p_n) \mathbf{A}^H) \geq \text{trace} [\mathbf{A} \text{diag}(\log p_1, \dots, \log p_n) \mathbf{A}^H] \quad (71)$$

if  $p_1, \dots, p_n > 0$ .

With Lemma 1, we can lower-bound (69) such that

$$\begin{aligned} & \mathbb{E}_{\mathbf{x}} \left[ \log \det \left( \frac{\sigma_h^2}{2f_d\sigma_n^2} \tilde{\mathbf{F}}^H \mathbf{X}^H \mathbf{X} \tilde{\mathbf{F}} + \mathbf{I}_{2[f_dN]+1} \right) \right] \\ & \geq \mathbb{E}_{\mathbf{x}} \left[ \text{trace} \left[ \tilde{\mathbf{F}}^H \text{diag} \left( \log \left( \frac{\sigma_h^2 |x_1|^2}{2f_d\sigma_n^2} + 1 \right), \dots, \log \left( \frac{\sigma_h^2 |x_N|^2}{2f_d\sigma_n^2} + 1 \right) \right) \tilde{\mathbf{F}} \right] \right] \\ & = \text{trace} \left[ \tilde{\mathbf{F}}^H \text{diag} \left( \mathbb{E}_x \log \left( \frac{\sigma_h^2 |x_1|^2}{2f_d\sigma_n^2} + 1 \right), \dots, \mathbb{E}_x \log \left( \frac{\sigma_h^2 |x_N|^2}{2f_d\sigma_n^2} + 1 \right) \right) \tilde{\mathbf{F}} \right] \\ & \stackrel{(a)}{=} \sum_{k=1}^{2[f_dN]+1} \mathbb{E}_x \log \left( \frac{\sigma_h^2}{2f_d\sigma_n^2} |x|^2 + 1 \right) \end{aligned} \quad (72)$$

where (a) results, because all  $x_k$  are identically distributed and because the columns of  $\tilde{\mathbf{F}}$  are orthonormal. Hence, with (68) a lower bound on the entropy rate is given by

$$\begin{aligned} h'(\mathbf{y}|\mathbf{x}) & \geq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^{2[f_dN]+1} \mathbb{E}_x \log \left( \frac{\sigma_h^2}{2f_d\sigma_n^2} |x|^2 + 1 \right) + \log(\pi e \sigma_n^2) \\ & = 2f_d \mathbb{E}_x \log \left( \frac{\sigma_h^2}{2f_d\sigma_n^2} |x|^2 + 1 \right) + \log(\pi e \sigma_n^2) = h'_L(\mathbf{y}|\mathbf{x}). \end{aligned} \quad (73)$$

Thus, we have found a lower bound on the entropy rate  $h'(\mathbf{y}|\mathbf{x})$  for identically distributed (i.d.) input distributions. To the best of our knowledge, this is the only known lower bound on the entropy rate  $h'(\mathbf{y}|\mathbf{x})$  which is not based on a peak power constraint.

For independently identically distributed (i.i.d.) zero-mean proper Gaussian input symbols the lower bound  $h'_L(\mathbf{y}|\mathbf{x})$  becomes

$$h'_L(\mathbf{y}|\mathbf{x}) = 2f_d \int_{z=0}^{\infty} \log \left( \frac{\sigma_h^2 \sigma_x^2}{2f_d\sigma_n^2} z + 1 \right) e^{-z} dz + \log(\pi e \sigma_n^2). \quad (74)$$

*Discussion on the Assumption of a Rectangular PSD:* For the case of constant modulus (CM) input distributions, it can be shown that the rectangular PSD maximizes  $h'(\mathbf{y}|\mathbf{x})$  among all PSDs with a compact support interval  $[-f_d, f_d]$  and a channel power  $\sigma_h^2$ . For the proof of this statement, we have to calculate  $\sup_{S_h(f) \in \mathcal{S}} h'(\mathbf{y}|\mathbf{x})|_{\text{CM}}$  where the set  $\mathcal{S}$  of PSDs is given by

$$\mathcal{S} = \left\{ S_h(f) = 0 \text{ for } f_d < |f| \leq 0.5, \int_{-\frac{1}{2}}^{\frac{1}{2}} S_h(f) df = \sigma_h^2 \right\}. \quad (75)$$

With (45) and (46), we get

$$\begin{aligned} \sup_{S_h(f) \in \mathcal{S}} h'(\mathbf{y}|\mathbf{x})|_{\text{CM}} & = \sup_{S_h(f) \in \mathcal{S}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log (\pi e (S_h(f) \sigma_x^2 + \sigma_n^2)) df \\ & = \sup_{S_h(f) \in \mathcal{S}} \int_{-f_d}^{f_d} \log (\pi e (S_h(f) \sigma_x^2 + \sigma_n^2)) df \end{aligned} \quad (76)$$

$$\stackrel{(a)}{=} \int_{-f_d}^{f_d} \log \left( \pi e \left( \frac{\sigma_h^2}{2f_d} \sigma_x^2 + \sigma_n^2 \right) \right) df \quad (77)$$

i.e., the PSD  $S_h(f)$  which maximizes  $h'(\mathbf{y}|\mathbf{x})$  is rectangular

$$S_h(f) = \begin{cases} \frac{\sigma_h^2}{2f_d} & \text{for } |f| \leq f_d \\ 0 & \text{otherwise} \end{cases}. \quad (78)$$

The last step in (77) can be proven as follows. The argument of the supremum in (76) is concave on the convex set  $\mathcal{S}$ . To find the  $S_h(f)$  that maximizes the supremum in (76), we define the functional

$$J(S_h) = \int_{-f_d}^{f_d} \log(\pi e (S_h(f)\sigma_x^2 + \sigma_n^2)) df + c \left( \int_{-f_d}^{f_d} S_h(f) df - \sigma_h^2 \right) \quad (79)$$

where  $c$  is a constant and the last term accounts for the constraint

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} S_h(f) df = \sigma_h^2. \quad (80)$$

For the  $S_h(f)$  that maximizes, the following equation must be fulfilled for each  $f$  within the interval  $[-f_d, f_d]$

$$\frac{\partial J}{\partial S_h(f)} = \frac{\sigma_x^2}{S_h(f)\sigma_x^2 + \sigma_n^2} + c = 0. \quad (81)$$

As this equation has to be fulfilled for each  $f$  and constant  $c$ ,  $S_h(f)$  must be constant within the interval  $[-f_d, f_d]$ . As the second derivative of  $J$  with respect to  $S_h(f)$  is negative for all  $S_h(f)$  included in  $\mathcal{S}$ , the given extremum is a maximum. Thus, with (80), (77) follows.

We conjecture that a rectangular PSD of the channel fading process maximizes  $h'(\mathbf{y}|\mathbf{x})$  for any i.i.d. input distribution with an average power  $\sigma_x^2$ , including the case of i.i.d. zero-mean proper Gaussian input symbols. Concerning this discussion see also [26, Section IV-A]. Consequently, the lower bound in (73) then holds only for a rectangular PSD. As this lower bound on  $h'(\mathbf{y}|\mathbf{x})$  is finally used for the upper bound on  $\mathcal{I}'(\mathbf{y}; \mathbf{x})$ , following the preceding conjecture, we get an upper bound on the achievable rate for a given maximum Doppler spread  $f_d$  for the worst case PSD.

#### D. The Achievable Rate

Based on the upper and lower bounds on  $h'(\mathbf{y})$  and  $h'(\mathbf{y}|\mathbf{x})$ , we are now able to give upper and lower bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols.

1) *Lower Bound:* With (27), (33), and (45), we get the following lower bound on the capacity

$$\begin{aligned} \mathcal{I}'(\mathbf{y}; \mathbf{x}) &\geq h'_L(\mathbf{y}) - h'_U(\mathbf{y}|\mathbf{x}) \\ &= \int_{z=0}^{\infty} \log(1 + \rho z) e^{-z} dz - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(1 + \rho \frac{S_h(f)}{\sigma_h^2}\right) df \\ &= \mathcal{I}'_L(\mathbf{y}; \mathbf{x}) \end{aligned} \quad (82)$$

where  $\rho$  is the average SNR as defined in (11). Notice that lower bounds on the achievable rate are also lower bounds on the capacity. Therefore, in the context of this lower bound we use the term *capacity* in the following. The lower bound in (82) is achievable with i.i.d. zero-mean proper Gaussian input symbols. As already stated, the lower bound on the capacity given in (82) is already known from [16]. The bound in (82) holds for an arbitrary PSD of the channel fading process with compact support. For the special case of a rectangular PSD as given in (7) the lower bound in (82) becomes

$$\mathcal{I}'_L(\mathbf{y}; \mathbf{x})|_{\text{Rect}} = \int_0^{\infty} \log(\rho z + 1) e^{-z} dz - 2f_d \log\left(\frac{\rho}{2f_d} + 1\right). \quad (83)$$

As the mutual information rate is nonnegative, we can further modify the lower bound in (82) as follows:

$$\mathcal{I}'_{L_{mod}}(\mathbf{y}; \mathbf{x}) = \max\{\mathcal{I}'_L(\mathbf{y}; \mathbf{x}), 0\}. \quad (84)$$

2) *Upper Bound:* Using (27), (36), and (74) we can upper-bound the achievable rate with i.i.d. zero-mean proper Gaussian input symbols and a rectangular PSD of the channel fading process by

$$\begin{aligned}\mathcal{I}'(\mathbf{y}; \mathbf{x}) &\geq h'_U(\mathbf{y}) - h'_L(\mathbf{y}|\mathbf{x}) \\ &= \log(\rho + 1) - 2f_d \int_0^\infty \log\left(\frac{\rho}{2f_d}z + 1\right) e^{-z} dz \\ &= \mathcal{I}'_U(\mathbf{y}; \mathbf{x}).\end{aligned}\tag{85}$$

Notice that for the derivation of this upper bound the assumption on independent input symbols has not been used. On the other hand, the restriction to identically distributed input symbols is required for the calculation of  $h'_L(\mathbf{y}|\mathbf{x})$  and, thus, for  $\mathcal{I}'_U(\mathbf{y}; \mathbf{x})$ .

To the best of our knowledge, the upper bound in (85) is new. Most other available upper bounds on the capacity hold only for input distributions with a peak power constraint and become loose for high peak-to-average power ratios, see, e.g., [2] and [1]. However, it has to be stated that the peak power constrained upper bounds in [2] and [1] are upper bounds on capacity and hold for an arbitrary PSD of the channel fading process.

As the mutual information rate in case of perfect channel state information at the receiver  $\mathcal{I}'(\mathbf{x}; \mathbf{y}|\mathbf{h})$  always upper-bounds the mutual information rate in the absence of channel state information, i.e.,

$$\mathcal{I}'(\mathbf{y}; \mathbf{x}) \leq \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})\tag{86}$$

we can modify the upper bound in (85) as follows:

$$\mathcal{I}'_{U_{mod}}(\mathbf{y}; \mathbf{x}) = \min\{\mathcal{I}'_U(\mathbf{y}; \mathbf{x}), \mathcal{I}'(\mathbf{x}; \mathbf{y}|\mathbf{h})\}\tag{87}$$

with  $\mathcal{I}'(\mathbf{x}; \mathbf{y}|\mathbf{h})$  given in (31).

3) *Tightness of Bounds:* In the following we study the tightness of the given bounds on the achievable rate for i.i.d. zero-mean proper Gaussian input symbols.

Fig. 1 shows the upper bound (85)/(87) and the lower bound (83)/(84) on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols as a function of the channel dynamics, which is characterized by  $f_d$ , in case the PSD of the channel fading process is rectangular for different SNRs. Obviously, the achievable rate strongly decreases with increasing channel dynamics, i.e.,  $f_d$ . Furthermore, the gap between the upper and the lower bound depends on the SNR and gets larger with an increasing SNR. In the following, we study the tightness of the given bounds analytically. This examination will show that the gap between the upper and the lower bound is bounded.

To evaluate the tightness of the upper and the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, we first evaluate the tightness of the upper and the lower bound on the channel output entropy rate  $h'(\mathbf{y})$ . Afterwards, we evaluate the tightness of the upper and lower bound on  $h'(\mathbf{y}|\mathbf{x})$ .

The difference between the upper bound  $h'_U(\mathbf{y})$  and the lower bound  $h'_L(\mathbf{y})$  in (36) and (33) is given by

$$\begin{aligned}\Delta_{h'(\mathbf{y})} &= h'_U(\mathbf{y}) - h'_L(\mathbf{y}) \\ &= \log(1 + \rho) - \int_0^\infty \log(1 + \rho z) e^{-z} dz.\end{aligned}\tag{88}$$

Fig. 7 in the Appendix B shows this difference. For  $\rho \rightarrow 0$  the difference  $\Delta_{h'(\mathbf{y})}$  converges to zero. And for  $\rho \rightarrow \infty$  the difference is given by

$$\lim_{\rho \rightarrow \infty} \Delta_{h'(\mathbf{y})} = \gamma \approx 0.57721 \quad [\text{nat/cu}]\tag{89}$$

where  $\gamma$  is the Euler constant. The limit in (89) can be found in [10].



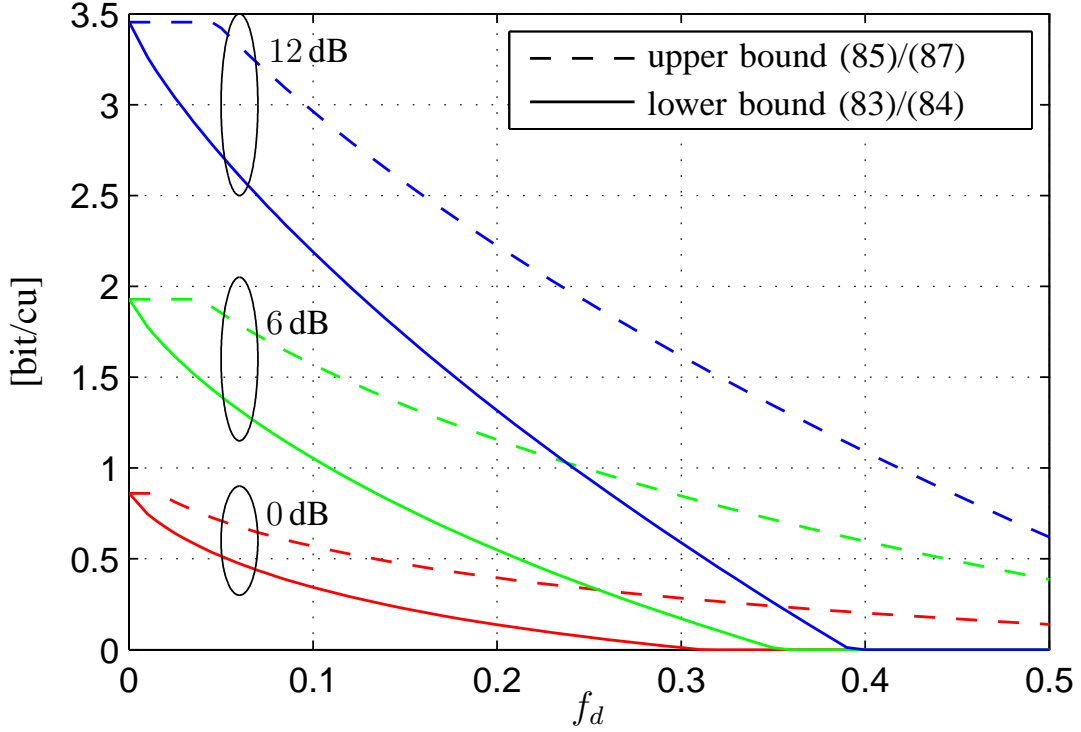


Fig. 1. Upper and lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian input distribution on a Rayleigh flat-fading channel with a rectangular PSD in bits per channel use (cu) over  $f_d$

The difference  $\Delta_{h'(\mathbf{y})}$  monotonically increases with the SNR, as

$$\begin{aligned} \frac{\partial \Delta_{h'(\mathbf{y})}}{\partial \rho} &= \frac{1}{1+\rho} - \int_0^\infty \frac{z}{1+\rho z} e^{-z} dz \\ &\stackrel{(a)}{\geq} \frac{1}{1+\rho} - \frac{1}{1+\rho} = 0 \end{aligned} \quad (90)$$

where for (a) we have used that  $\frac{z}{1+\rho z}$  is concave in  $z$  and, thus, we can apply Jensen's inequality. Thus,  $\Delta_{h'(\mathbf{y})}$  is bounded by

$$0 \leq \Delta_{h'(\mathbf{y})} \leq \gamma. \quad (91)$$

In addition, the difference between the upper bound and the lower bound on  $h'(\mathbf{y}|\mathbf{x})$  in case of a rectangular PSD is given by, cf. (45) and (74)

$$\begin{aligned} \Delta_{h'(\mathbf{y}|\mathbf{x})} &= h'_U(\mathbf{y}|\mathbf{x}) - h'_L(\mathbf{y}|\mathbf{x}) \\ &= 2f_d \left\{ \log \left( 1 + \frac{\rho}{2f_d} \right) - \int_0^\infty \log \left( 1 + \frac{\rho}{2f_d} z \right) e^{-z} dz \right\}. \end{aligned} \quad (92)$$

For asymptotically small Doppler frequencies  $\Delta_{h'(\mathbf{y}|\mathbf{x})}$  approaches zero independently of the SNR. Furthermore, observing the structural similarity between (92) and (88), it can be shown that

$$\lim_{\rho \rightarrow 0} \Delta_{h'(\mathbf{y}|\mathbf{x})} = 0 \quad (93)$$

independently of  $f_d$ . For asymptotically high SNR and a fixed  $f_d$  the difference is bounded by

$$\lim_{\rho \rightarrow \infty} \Delta_{h'(\mathbf{y}|\mathbf{x})} = 2f_d \gamma \approx 2f_d \cdot 0.57721 \quad [\text{nat/cu}] \quad (94)$$

where the same limit as in (89) is used. Corresponding to  $\Delta_{h'(\mathbf{y})}$ ,  $\Delta_{h'(\mathbf{y}|\mathbf{x})}$  is monotonically increasing with the SNR and thus, it can be bounded by

$$0 \leq \Delta_{h'(\mathbf{y}|\mathbf{x})} \leq \gamma 2f_d \quad [\text{nat/cu}]. \quad (95)$$

With  $\Delta_{h'(\mathbf{y})}$  and  $\Delta_{h'(\mathbf{y}|\mathbf{x})}$ , the difference between the upper bound  $\mathcal{I}'_U(\mathbf{y}; \mathbf{x})$  and the lower bound  $\mathcal{I}'_L(\mathbf{y}; \mathbf{x})$  for i.i.d. zero-mean proper Gaussian input symbols and a rectangular PSD is given by

$$\begin{aligned} \Delta_{\mathcal{I}'(\mathbf{y}; \mathbf{x})} &= \mathcal{I}'_U(\mathbf{y}; \mathbf{x}) - \mathcal{I}'_L(\mathbf{y}; \mathbf{x})|_{\text{Rect}} \\ &= \Delta_{h'(\mathbf{y})} + \Delta_{h'(\mathbf{y}|\mathbf{x})}. \end{aligned} \quad (96)$$

Hence, we get the following limits

$$\begin{aligned} \lim_{\rho \rightarrow 0} \Delta_{\mathcal{I}'(\mathbf{y}; \mathbf{x})} &= 0 \\ \lim_{\rho \rightarrow \infty} \Delta_{\mathcal{I}'(\mathbf{y}; \mathbf{x})} &= \gamma(1 + 2f_d) \end{aligned} \quad (97)$$

and as  $\Delta_{h'(\mathbf{y})}$ ,  $\Delta_{h'(\mathbf{y}|\mathbf{x})}$ , and, thus,  $\Delta_{\mathcal{I}'(\mathbf{y}; \mathbf{x})}$  monotonically increase with the SNR, we can bound the difference by

$$0 \leq \Delta_{\mathcal{I}'(\mathbf{y}; \mathbf{x})} \leq \gamma(1 + 2f_d) \quad [\text{nat/cu}]. \quad (98)$$

*a) Asymptotically Small Channel Dynamics:* For asymptotically small channel dynamics, i.e.,  $f_d \rightarrow 0$ , the lower bound  $\mathcal{I}'_L(\mathbf{y}; \mathbf{x})$  in (82) converges to the mutual information rate in case of perfect channel knowledge in (31)

$$\lim_{f_d \rightarrow 0} \mathcal{I}'_L(\mathbf{y}; \mathbf{x}) = \mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h}) \quad (99)$$

i.e., to the coherent capacity. This corresponds to the physical interpretation that a channel that changes arbitrarily slowly can be estimated arbitrarily well, and, therefore, the penalty term  $\mathcal{I}'(\mathbf{x}; \mathbf{h}|\mathbf{y})$  in (24) approaches zero. Thus, for  $f_d \rightarrow 0$ , the lower bound  $\mathcal{I}'_L(\mathbf{y}; \mathbf{x})$  is tight.

### E. The Asymptotic High SNR Behavior

In this section, we examine the slope of the achievable rate over the SNR for asymptotically large SNRs depending on the channel dynamics. For a compactly supported PSD the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given in (82) is characterized by the following high SNR slope<sup>4</sup>, which is often named *pre-log*

$$\begin{aligned} \lim_{\rho \rightarrow \infty} \frac{\partial \mathcal{I}'_L(\mathbf{y}; \mathbf{x})}{\partial \log(\rho)} &= \lim_{\rho \rightarrow \infty} \frac{\partial}{\partial \log(\rho)} \left[ \int_{z=0}^{\infty} \log(\rho z + 1) e^{-z} dz - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left( \frac{S_h(f)}{\sigma_h^2} \rho + 1 \right) df \right] \\ &= \lim_{\rho \rightarrow \infty} \left[ \int_{z=0}^{\infty} \frac{\rho z}{\rho z + 1} e^{-z} dz - \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\frac{S_h(f)}{\sigma_h^2} \rho}{\frac{S_h(f)}{\sigma_h^2} \rho + 1} df \right] \\ &= 1 - 2f_d \end{aligned} \quad (100)$$

as  $S_h(f) \neq 0$  for  $|f| \leq f_d$ .

The upper bound  $\mathcal{I}'_U(\mathbf{y}; \mathbf{x})$  holds only for the special case of a rectangular PSD of the channel fading process. For this case the difference between the upper bound  $\mathcal{I}'_U(\mathbf{y}; \mathbf{x})$  and the lower bound  $\mathcal{I}'_L(\mathbf{y}; \mathbf{x})$  converges to a constant for high SNR, cf. (97). Thus, both bounds must have the same asymptotic high SNR slope and we conjecture that the achievable rate  $\mathcal{I}'(\mathbf{y}; \mathbf{x})$  is also characterized by the same asymptotic

<sup>4</sup>When using the term *high SNR slope* we refer to the high SNR limit of the derivative of the achievable rate (bound) with respect to the logarithm of the SNR.

SNR slope. In [3], it has been shown that the high SNR slope (pre-log) of the peak power constrained capacity also corresponds to  $1 - 2f_d$ . For a more detailed discussion on this, we refer to Section III-F.

It is interesting to note that the high SNR slope of the achievable rate is degraded by the term  $2f_d$ . Now, recall the discussion on the limitations of the discrete-time input-output relation in Section II-B. There it has been shown that symbol rate sampling does not yield a signal representation with a sufficient statistic, as the normalized received signal bandwidth is given by  $1 + 2f_d$  (for a roll-off factor  $\beta_{ro} = 0$ ). The excess bandwidth leading to aliasing is given by  $2f_d$ , which exactly corresponds to the degradation of the high SNR slope of the achievable rate. Up to now, we do not know, if there is an implicit relation between these observations.

#### F. Comparison to Asymptotes in [3]

In [3], Lapidoth gives bounds for the capacity of noncoherent Rayleigh fading channels. These bounds are mainly derived to evaluate the asymptotic high SNR behavior. He distinguishes between two cases, *nonregular* and *regular* fading introduced by Doob [14]. The case of *nonregular* fading is characterized by the property that the prediction error variance of a one-step channel predictor — having infinitely many observations in the past — asymptotically approaches zero, when the SNR approaches infinity. As we consider the case that the PSD of the channel fading process is bandlimited with  $f_d < 0.5$ , our scenario corresponds to the *nonregular* case in [3], which is also named *pre-log* case. In contrast to our bounds on the achievable rate, where we assume i.i.d. zero-mean proper Gaussian input symbols with an average power  $\sigma_x^2$ , [3] does not constrain the input distribution except of a peak power constraint.

The capacity bounds in [3] are given by [3, eq. (33) and (47)]

$$C \leq \log \log \tilde{\rho} - \gamma - 1 + \log \left( \frac{1}{\epsilon_{pred}^2(1/\tilde{\rho})} \right) + o(1) \quad (101)$$

$$C \geq \log \left( \frac{1}{\epsilon_{pred}^2(4/\tilde{\rho}) + \frac{8}{5\tilde{\rho}}} \right) - \gamma - \log \left( \frac{1}{1 - \epsilon_{pred}^2(4/\tilde{\rho})} \right) - \log \left( \frac{5e}{6} \right), \quad (102)$$

where  $\gamma \approx 0.577$  is the Euler constant and  $\tilde{\rho}$  is defined as

$$\tilde{\rho} = \frac{P_{\text{peak}} \sigma_h^2}{\sigma_n^2} \quad (103)$$

i.e., it is an alternative definition of an SNR based on the peak power  $P_{\text{peak}}$  instead of the average power  $\sigma_x^2$  used for the definition of the average SNR  $\rho$ . Furthermore,  $o(1)$  depends on the SNR and converges to zero for  $\tilde{\rho} \rightarrow \infty$ , i.e.,  $f(n) \in o(g(n))$  if

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0. \quad (104)$$

In addition, the prediction error variance  $\epsilon_{pred}^2(\delta^2)$  is given by

$$\epsilon_{pred}^2(\delta^2) = \exp \left( \int_{-0.5}^{0.5} \log \left( \frac{S_h(f)}{\sigma_h^2} + \delta^2 \right) df \right) - \delta^2. \quad (105)$$

Although for the bounds on the peak power constrained capacity in [3] not an explicit average power constraint has been used, but only a peak power constraint, by this peak power constraint implicitly also a constraint on the average power is given. This should be obvious, as for the average power  $\sigma_x^2$  the inequality  $\sigma_x^2 \leq P_{\text{peak}}$  must hold. Furthermore, it has to be considered that in case of using a peak power constraint, it is in general not optimal to use the maximum average power  $\sigma_x^2$ . For a discussion on this see below in Section III-G1. In case the maximum average power  $\sigma_x^2$  is not used, i.e.,  $\mathbb{E}[|x_k|^2] < \sigma_x^2$ , the SNR  $\rho$  as defined in (11) is not the actual average SNR. Therefore, in the case of using a peak power constraint,  $\rho$  is named *nominal* average SNR. However, in the case of i.i.d. zero-mean proper Gaussian

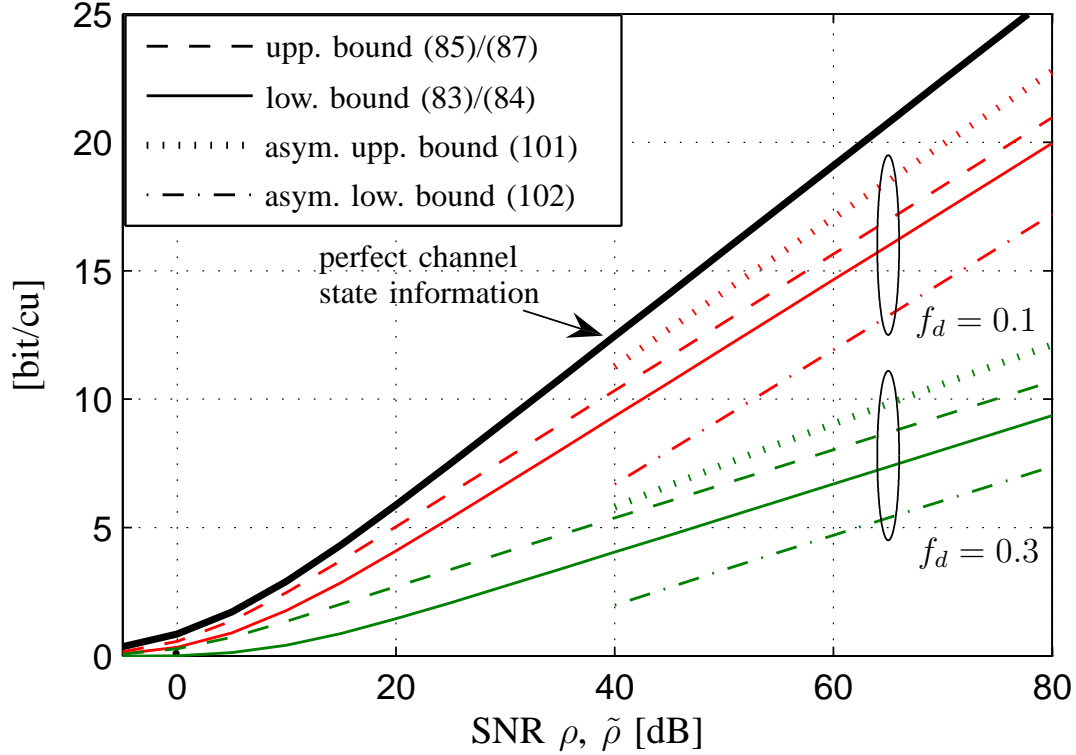


Fig. 2. Comparison of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian inputs in (85)/(87) and (83)/(84) (SNR  $\rho$ ) with asymptotic bounds on the peak power constrained capacity in (101) and (102) (SNR  $\tilde{\rho}$ ), [3, eq. (33) and (47)] (The asymptotic upper bound (101) only holds for  $\tilde{\rho} \rightarrow \infty$  as we neglect the term  $o(1)$  in (101), which approaches zero for  $\tilde{\rho} \rightarrow \infty$ ); rectangular PSD of the channel fading process

input symbols, the achievable rate is maximized when using the maximum average power  $\sigma_x^2$ , i.e., in this case the nominal average SNR  $\rho$  is also the actual average SNR.

As the peak power constraint that has been used for the bounds on the peak power constrained capacity in [3], i.e., for (101) and (102), implicitly constrains the average power to  $\sigma_x^2$ , for the comparison of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols and the bounds on the peak power constrained capacity in [3], we choose  $\tilde{\rho}$  in (101) and (102) to be equal to the nominal average SNR  $\rho$  used for the bounds on the achievable rate with i.i.d. Gaussian input symbols, i.e., set  $\sigma_x^2 = P_{\text{peak}}$ .

Fig. 2 shows a comparison of the lower bound on the capacity in (83)/(84) and the upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs in (85)/(87) with the high SNR asymptotes for the capacity in the corresponding pre-log case given in [3], i.e., (101) and (102). The bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, i.e., the lower bound in (83)/(84) and our upper bound in (85)/(87), are in between the asymptotes for the upper bound and the lower bound on capacity given in [3]. However, the bounds in [3] consider a peak power constrained input distribution. Therefore, this comparison is not absolutely fair. In addition, and this is the main observation from this comparison, our bounds have the same slope in the high SNR regime as the high SNR asymptotes for the peak power constrained capacity in [3].

#### G. Comparison to Capacity Bounds for Peak Power Constrained Inputs in [2] and [1]

In the following, we will draw the connection of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given in (82)/(84) and (85)/(87) with the bounds on the peak power

constrained capacity given in [2] and [1]. The peak power constrained capacity is defined by

$$C_{\text{peak}} = \lim_{N \rightarrow \infty} \sup_{\mathcal{P}^{\text{peak}}} \frac{1}{N} \mathcal{I}(\mathbf{y}; \mathbf{x}) \quad (106)$$

with  $\mathcal{P}^{\text{peak}}$  being the set of peak power constrained probability density functions of the input distribution given by

$$\mathcal{P}^{\text{peak}} = \left\{ p(\mathbf{x}) \left| \mathbf{x} \in \mathbb{C}^N, \frac{1}{N} \mathbb{E} [\mathbf{x}^H \mathbf{x}] \leq \sigma_x^2, |x_k|^2 \leq P_{\text{peak}} \forall k \right. \right\}. \quad (107)$$

Following along the lines of the derivation of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, we discuss the differences when deriving bounds on the capacity with peak power constrained input symbols.

1) *Upper Bound:* We start with the derivation of the upper bound on the peak power constrained capacity. With (27), an upper bound on the peak power constrained capacity can be given based on an upper bound on the output entropy  $h'(\mathbf{y})$  and a lower bound on the conditional output entropy  $h'(\mathbf{y}|\mathbf{x})$  resulting in<sup>5</sup>

$$\sup_{\mathcal{P}^{\text{peak}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) \leq \sup_{\mathcal{P}^{\text{peak}}} \{h'_U(\mathbf{y}) - h'_L(\mathbf{y}|\mathbf{x})\}. \quad (108)$$

Note that at the moment  $h'_U(\mathbf{y})$  and  $h'_L(\mathbf{y}|\mathbf{x})$  are only place holders for upper and lower bounds, which are not yet further specified. In the following, we will relate the derivation of the corresponding bounds given above for i.i.d. zero-mean proper Gaussian input symbols to the case of peak power constrained input symbols considered here.

As the derivation of the lower bound on  $h'(\mathbf{y}|\mathbf{x})$  in Section III-C2 relies on the assumption on identically distributed (i.d.) input symbols, in a first step we restrict to this kind of input distributions and, therefore, define the following set of probability density functions:

$$\mathcal{P}_{\text{i.d.}}^{\text{peak}} = \left\{ p(\mathbf{x}) \left| \mathbf{x} \in \mathbb{C}^N, p(x_i) = p(x_j) \quad \forall i, j, \{ \mathbb{E}[|x_k|^2] \leq \sigma_x^2, |x_k|^2 \leq P_{\text{peak}} \} \forall k \right. \right\} \quad (109)$$

which corresponds to the set  $\mathcal{P}^{\text{peak}}$  in (107) with the further restriction that the input symbols are identically distributed. I.e., we derive an upper bound on

$$\begin{aligned} \sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) &\leq \sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \{h'_U(\mathbf{y}) - h'_L(\mathbf{y}|\mathbf{x})\} \\ &= \sup_{\alpha \in [0,1]} \sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}} | \alpha} \{h'_U(\mathbf{y}) - h'_L(\mathbf{y}|\mathbf{x})\}. \end{aligned} \quad (110)$$

The calculation of the supremum in (110) is done in two steps. The inner supremum is taken over the constrained set  $\mathcal{P}_{\text{i.d.}}^{\text{peak}} | \alpha$  being characterized by an average power  $\alpha \sigma_x^2$  which holds with equality. Because of the fact that in (109) we only use a constraint on the maximum average input power given by  $\sigma_x^2$  the outer supremum is taken over  $\alpha \in [0, 1]$ . The set  $\mathcal{P}_{\text{i.d.}}^{\text{peak}} | \alpha$  is given by

$$\mathcal{P}_{\text{i.d.}}^{\text{peak}} | \alpha = \left\{ p(\mathbf{x}) \left| \mathbf{x} \in \mathbb{C}^N, p(x_i) = p(x_j) \quad \forall i, j, \{ \mathbb{E}[|x_k|^2] = \alpha \sigma_x^2, |x_k|^2 \leq P_{\text{peak}} \} \forall k \right. \right\} \quad (111)$$

which corresponds to the set  $\mathcal{P}_{\text{i.d.}}^{\text{peak}}$  except that the average power is now fixed to  $\alpha \sigma_x^2$  with equality. Such a separation has also been used in [1] and in [8].

<sup>5</sup>Note that in (108) we make a slight misuse of notation. The set  $\mathcal{P}^{\text{peak}}$  is defined for input vectors  $\mathbf{x}$  of length  $N$ . Therefore, the exchange of the limit and the supremum as it is used in (108) while using the mutual information rate is formally not correct. However, to avoid a further complication of notation, we use the set  $\mathcal{P}^{\text{peak}}$  also in the context of information rates. The same holds also in the following for other sets of input distributions.

For the evaluation of (110) we require an upper bound on  $h'(\mathbf{y})$  and a lower bound on  $h'(\mathbf{y}|\mathbf{x})$  which hold for all input distributions contained in the set  $\mathcal{P}_{\text{i.d.}}^{\text{peak}}|_{\alpha}$ .

All steps of the derivation of the upper bound on  $h'(\mathbf{y})$  in Section III-B2 for the case i.i.d. zero-mean proper Gaussian input symbols hold also for all input distributions contained in the set  $\mathcal{P}_{\text{i.d.}}^{\text{peak}}|_{\alpha}$ , except that (a) in (34) is now an inequality (Hadamard's inequality), as the matrix  $\mathbf{R}_y$  is not necessarily diagonal as we have dropped the restriction to i.i.d. input symbols. Furthermore, the average transmit power of the input symbols is now fixed to  $\alpha\sigma_x^2$  and, thus, we get the following upper bound:

$$h'_U(\mathbf{y})|_{p(\mathbf{x}) \in \mathcal{P}_{\text{i.d.}}^{\text{peak}}|_{\alpha}} = \log(\pi e(\alpha\sigma_x^2\sigma_h^2 + \sigma_n^2)). \quad (112)$$

Concerning the lower bound on  $h'(\mathbf{y}|\mathbf{x})$  up to (73) only the assumption on i.i.d. input symbols has been used, which also holds for all input probability density functions contained in  $\mathcal{P}_{\text{i.d.}}^{\text{peak}}|_{\alpha}$ . Thus, substituting (112) and (73) into (110) we get

$$\begin{aligned} \sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) &\leq \sup_{\alpha \in [0,1]} \sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}|_{\alpha}} \{h'_U(\mathbf{y}) - h'_L(\mathbf{y}|\mathbf{x})\} \\ &= \sup_{\alpha \in [0,1]} \sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}|_{\alpha}} \left\{ \log(\alpha\rho + 1) - 2f_d E_x \log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}|x|^2 + 1\right) \right\} \\ &= \sup_{\alpha \in [0,1]} \left\{ \log(\alpha\rho + 1) - 2f_d \inf_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}|_{\alpha}} E_x \log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}|x|^2 + 1\right) \right\} \end{aligned} \quad (113)$$

with the nominal average SNR  $\rho$  given in (11).

The term containing the infimum on the RHS of (113) can be lower-bounded in the following way:

$$\begin{aligned} \inf_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}|_{\alpha}} E_x \log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}|x|^2 + 1\right) &= \inf_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}|_{\alpha}} \int_{|x|=0}^{\sqrt{P_{\text{peak}}}} \frac{\log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}|x|^2 + 1\right)}{|x|^2} |x|^2 p(|x|) d|x| \\ &\stackrel{(a)}{\geq} \frac{\log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}P_{\text{peak}} + 1\right)}{P_{\text{peak}}} \inf_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}|_{\alpha}} \int_{|x|=0}^{\sqrt{P_{\text{peak}}}} |x|^2 p(|x|) d|x| \\ &= \frac{\log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}P_{\text{peak}} + 1\right)}{P_{\text{peak}}} \alpha\sigma_x^2 \end{aligned} \quad (114)$$

where for (a) we have used that all factors of the integrand are positive and that the term

$$\frac{\log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2}|x|^2 + 1\right)}{|x|^2} = \frac{1}{z} \log(cz + 1) \quad (115)$$

with  $c = \frac{\sigma_h^2}{2f_d\sigma_n^2}$  and  $z = |x|^2$  is monotonically decreasing in  $z$  as

$$\begin{aligned} \frac{\partial}{\partial z} \left\{ \frac{1}{z} \log(cz + 1) \right\} &= \frac{c}{(cz + 1)z} - \frac{\log(cz + 1)}{z^2} < 0 \\ &\Leftrightarrow \frac{cz}{cz + 1} < \log(cz + 1) \end{aligned} \quad (116)$$

which holds for  $cz > -1$ . Thus, the term in (115) is minimized for  $z = |x|^2 = P_{\text{peak}}$ . A similar approach to calculate the infimum in (114) has been used in [27] for an analogous problem. Notice that the result given in (114) means that the infimum on  $h'_L(\mathbf{y}|\mathbf{x})$  for a fixed average transmit power is achieved with on-off keying.

With (114), we get the following upper bound on the RHS of (113):

$$\begin{aligned} \sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) &\leq \sup_{\alpha \in [0,1]} \left\{ \log(\alpha\rho + 1) - 2f_d \frac{\alpha\sigma_x^2}{P_{\text{peak}}} \log\left(\frac{\sigma_h^2}{2f_d\sigma_n^2} P_{\text{peak}} + 1\right) \right\} \\ &= \sup_{\alpha \in [0,1]} \left\{ \log(\alpha\rho + 1) - 2f_d \frac{\alpha}{\beta} \log\left(\frac{\rho\beta}{2f_d} + 1\right) \right\} \end{aligned} \quad (117)$$

with the *nominal peak-to-average power ratio*<sup>6</sup>

$$\beta = \frac{P_{\text{peak}}}{\sigma_x^2}. \quad (118)$$

As the argument of the supremum on the RHS of (117) is concave in  $\alpha$  and, thus, there exists a unique maximum, it can easily be shown that the supremum of (117) with respect to  $\alpha \in [0, 1]$  is given by

$$\alpha_{\text{opt}} = \min \left\{ 1, \left( \frac{2f_d}{\beta} \log\left(\frac{\rho\beta}{2f_d} + 1\right) \right)^{-1} - \frac{1}{\rho} \right\} \quad (119)$$

and, thus,

$$\begin{aligned} \sup_{\mathcal{P}_{\text{i.d.}}^{\text{peak}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) &\leq \log(\alpha_{\text{opt}}\rho + 1) - 2f_d \frac{\alpha_{\text{opt}}}{\beta} \log\left(\frac{\rho\beta}{2f_d} + 1\right) \\ &= \mathcal{I}'_U(\mathbf{y}; \mathbf{x})|_{P_{\text{peak}}}. \end{aligned} \quad (120)$$

With (120), we have found an upper bound on the achievable rate with i.d. input symbols and a peak power constraint for the special case of a rectangular PSD of the channel fading process. Note that the writing  $\mathcal{I}'_U(\mathbf{y}; \mathbf{x})|_{P_{\text{peak}}}$  denotes an upper bound on the peak power constrained achievable rate.

Note that  $\alpha_{\text{opt}} < 1$  corresponds to the case that it is not optimal to use the maximum average transmit power allowed by the set  $\mathcal{P}_{\text{i.d.}}^{\text{peak}}$ . This behavior is a result of the peak power constraint. Therefore, consider the extreme case  $\beta = 1$  and  $f_d = 0.5$ , i.e., an uncorrelated channel.  $\alpha = 1$  then would correspond to constant modulus signaling, i.e., the transmitter puts all information into the phase of the transmitted signal. As the channel is uncorrelated from symbol to symbol and unknown to the receiver, the mutual information rate  $\mathcal{I}'(\mathbf{y}; \mathbf{x})$  is zero. Therefore, it is better, if the receiver does not use all its transmit power, i.e., uses an  $\alpha < 1$ , enabling modulation of the magnitude, which leads to a positive  $\mathcal{I}'(\mathbf{y}; \mathbf{x})$ .

The choice  $\alpha_{\text{opt}} = 1$ , corresponding to the case that it is optimal to use the maximum possible average transmit power, can be shown to be optimal, on the one hand, if

$$1 \leq \rho \leq \frac{2f_d}{\beta} \left[ \exp\left(\frac{1}{2} \frac{\beta}{2f_d}\right) - 1 \right] \quad (121)$$

or, on the other hand, if

$$2f_d \leq \frac{\beta}{\rho + 2} \text{ for } \rho \leq 1. \quad (122)$$

<sup>6</sup>Instead of the common term *peak-to-average power ratio* we choose the term *nominal peak-to-average power ratio*, as in case of a peak power constraint it is not necessarily optimal to use the maximum average power  $\sigma_x^2$ . In case the actual average power is equal to the maximum average power  $\sigma_x^2$ ,  $\beta$  corresponds to the actual peak-to-average power ratio.

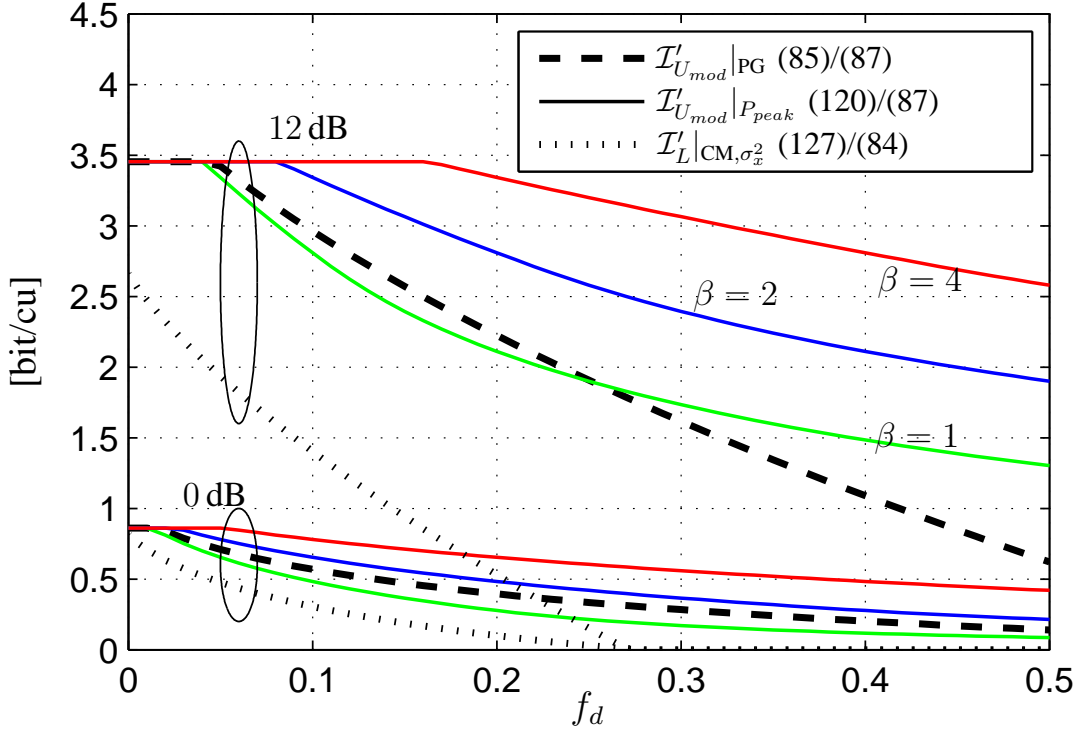


Fig. 3. Comparison of the upper bounds on the achievable rate for i.d. input symbols with a peak power constraint in (120)/(87) and with i.i.d. zero-mean proper Gaussian inputs (PG) in (85)/(87); (Note, (85) also holds for i.d. zero-mean proper Gaussian input symbols); in addition, the capacity lower bound (127)/(84) is shown, which is achievable with i.i.d. input symbols and a peak power constraint

For a proof of these conditions see Appendix C. As in realistic scenarios  $f_d$  is close to zero, the conditions (121) and (122) are typically fulfilled. However, for the parameter range displayed in Fig. 3 the conditions in (121) and (122) are not always fulfilled.<sup>7</sup>

In terms of the analytical expression, the upper bound on the achievable rate with i.d. input symbols in (120) is equal to the upper bound on the peak power constrained capacity given in [1, Prop. 2.2]. However, the upper bound in [1, Prop. 2.2] is, on the one hand, an upper bound on capacity as, except of the peak and average power constraints, no further assumptions on the input distributions have been made. On the other hand, the upper bound in [1, Prop. 2.2] holds for arbitrary PSDs of the channel fading process, while the derivation of the upper bound in (120) is based on the assumption of a rectangular PSD of the channel fading process. However, the approach of the derivation of the upper bound on the capacity given in [1, Prop. 2.2] is completely different to our approach and is inherently based on the peak power constraint, while we use this peak power constraint only in the last step of the derivation.

<sup>7</sup>Note that in case of i.i.d. zero-mean proper Gaussian input symbols as discussed before, it is not necessary to use the factor  $\alpha$ , i.e., consider cases where the actual average transmit power is smaller than the maximum average transmit power, as it can be shown that the upper bound in (85) is always maximized while using the maximum available average transmit power. The proof is based on the fact that (85) monotonically increases with  $\rho$  as

$$\begin{aligned} \frac{\partial}{\partial \rho} \left\{ \log(\rho + 1) - 2f_d \int_0^\infty \log\left(\frac{\rho}{2f_d}z + 1\right) e^{-z} dz \right\} &= \frac{1}{\rho + 1} - 2f_d \int_0^\infty \frac{\frac{1}{2f_d}z}{\frac{\rho}{2f_d}z + 1} e^{-z} dz \\ &\stackrel{(b)}{\geq} \frac{1}{\rho + 1} - 2f_d \frac{\frac{1}{2f_d}}{\frac{\rho}{2f_d} + 1} \geq 0 \end{aligned} \quad (123)$$

where for (b) we use that  $\frac{\frac{1}{2f_d}z}{\frac{\rho}{2f_d}z + 1}$  is concave in  $z$  and, thus, we can apply Jensen's inequality. This indicates that in case of the lack of a peak power constraint it is optimal to use the maximum average transmit power  $\sigma_x^2$ .



Therefore, our lower bound on  $h'(\mathbf{y}|\mathbf{x})$  in (73) also enables to give an upper bound on the achievable rate for non-peak power constrained input symbols like proper Gaussian input symbols.

As stated, we made the assumption on identically distributed (i.d.) input symbols in the derivation of our upper bound. We do not know if this assumption poses a real restriction in the sense of excluding the capacity-achieving input distribution. Therefore, it would be necessary to know if the capacity-achieving input distribution is characterized by identically distributed input symbols. We have no answer to this question. However, as in case of a peak power constraint our upper bound on the achievable rate given in (120) corresponds to the upper bound on the peak power constrained capacity given in [1, Prop. 2.2], the restriction to identically distributed inputs seems not to be a severe restriction in the sense that it leads to an upper bound being lower than the capacity.

However, in [1] it is shown that i.i.d. inputs, i.e., with an additional constraint on independent input symbols, are not capacity achieving in general. Based on the parameter

$$\lambda = \int_{-\frac{1}{2}}^{\frac{1}{2}} |S_h(f)|^2 df \quad (124)$$

it has been shown in [1] that under the assumption of an absolutely summable autocorrelation function  $r_h(l)$ , see (4), in the asymptotic low SNR limit i.i.d. inputs are only capacity-achieving in the following two cases

- if  $\lambda = \sigma_h^4$ , corresponding to a memoryless channel,
- or with a nominal peak-to-average power ratio of  $\beta = 1$  and  $\lambda \geq 2\sigma_h^4$ , i.e., when the fading process is *nonephemeral*.

Notice that the proof in [1] is explicitly based on the asymptotic low SNR limit. On the other hand, for the high SNR case i.i.d. zero-mean proper Gaussian inputs achieve the same asymptotic high SNR behavior, in terms of the slope (pre-log), as the peak power constrained channel capacity, as it has been discussed in Section III-F.

In Fig. 3, the upper bound on the achievable rate with a peak power constraint in (120) is shown for different nominal peak-to-average power ratios  $\beta$  in comparison to the upper bound on the achievable rate for zero-mean proper Gaussian input symbols in (85) (both combined with (87)). This comparison shows that except for  $\beta$  close to 1 and a small to average SNR or sufficiently small channel dynamics the upper bound on the achievable rate for proper Gaussian inputs is lower than the upper bound for peak power constrained input symbols in (120).

*a) High Nominal Peak-to-Average Power Ratios:* Considering higher order modulation, the nominal peak-to-average power ratio  $\beta$  may become relatively large. For proper Gaussian inputs it is in fact infinite. Obviously, for large peak powers  $P_{\text{peak}}$ , the second term in the upper bound on the RHS of (120) approaches zero and, thus

$$\lim_{\beta \rightarrow \infty} \mathcal{I}'_U(\mathbf{y}; \mathbf{x})|_{P_{\text{peak}}} = \log(\rho + 1) \quad (125)$$

which obviously is loose as this is the capacity of an AWGN channel being already larger than the coherent capacity of the fading channel. This underlines the value of the upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, which are not peak power limited and serve well to upper-bound the achievable rate with practical modulation and coding schemes.

It can be shown that the upper bound in (120) becomes loose for  $\beta > 1$  and high SNR. Therefore, we calculate the asymptotic high SNR slope of the peak power constrained upper bound given in (120), where for the moment we restrict to the case of using the maximum average power, i.e.,  $\alpha = 1$ , although this is in general not an upper bound on the achievable rate. The motivation for this will become obvious afterwards. For the peak power constrained upper bound given in (117) and for the special case  $\alpha = 1$

the derivative with respect to  $\log(\rho)$  in the high SNR limit is given by

$$\begin{aligned}
\lim_{\rho \rightarrow \infty} \frac{\partial \mathcal{I}'_U(\mathbf{y}; \mathbf{x})|_{P_{\text{peak}}, \alpha=1}}{\partial \log(\rho)} &= \lim_{\rho \rightarrow \infty} \frac{\partial}{\partial \log(\rho)} \left[ \log(\rho + 1) - \frac{2f_d}{\beta} \cdot \log\left(\frac{\rho\beta}{2f_d} + 1\right) \right] \\
&= \lim_{\rho \rightarrow \infty} \left[ \frac{\rho}{\rho + 1} - \frac{2f_d}{\beta} \frac{\frac{\beta}{2f_d}\rho}{\frac{\beta}{2f_d}\rho + 1} \right] \\
&= 1 - \frac{2f_d}{\beta}.
\end{aligned} \tag{126}$$

Obviously, if the nominal peak-to-average power ratio  $\beta$  is not equal to one, the slope of the peak power constrained upper bound with the constraint  $\alpha = 1$  is higher than the slope of the high SNR asymptote on the peak power constraint capacity given in [3], see (101), which is given by  $1 - 2f_d$ . The asymptotic bound in (101) holds for an arbitrary nominal peak-to-average power ratio  $\beta$ . As an optimized  $\alpha$  will lead to a larger upper bound, this unveils that the peak power constrained upper bound on the achievable rate in (120) is loose for  $\beta > 1$  and high SNR.

2) *Lower Bound:* As done before for the case of the upper bound, now we discuss the relation between the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols in (82) and the lower bound on the capacity for peak power constrained input symbols given in [2, (34)]. The lower bound on the achievable rate given in (82) obviously does not hold in case of a peak power constrained input, as in this case the coherent mutual information rate  $\mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$ , being used in (30) to calculate a lower bound on  $h'(\mathbf{y})$ , is smaller than (31), which holds for the case of i.i.d. zero-mean proper Gaussian inputs, the capacity-achieving input distribution in the coherent case.

As the mutual information for an arbitrary input distribution in the set  $\mathcal{P}^{\text{peak}}$  defined in (107) is a lower bound on the capacity with a peak power constrained input distribution, in a first step we assume a constant modulus (CM) input distribution. I.e., all input symbols have power  $\sigma_x^2$  and a uniformly distributed phase.

Based on (27) and (30), a lower bound on the mutual information rate with constant modulus input symbols is given by

$$\begin{aligned}
\sup_{\mathcal{P}^{\text{peak}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) &\geq \sup_{\mathcal{P}^{\text{peak}}} \{ \mathcal{I}(y; x|h) + h'(\mathbf{y}|\mathbf{x}, \mathbf{h}) - h'(\mathbf{y}|\mathbf{x}) \} \\
&\geq \{ \mathcal{I}(y; x|h) + h'(\mathbf{y}|\mathbf{x}, \mathbf{h}) - h'(\mathbf{y}|\mathbf{x}) \} |_{\text{CM}, \sigma_x^2} \\
&\stackrel{(a)}{=} \mathcal{I}(y; x|h) |_{\text{CM}, \sigma_x^2} - \int_{f=-\frac{1}{2}}^{\frac{1}{2}} \log \left( \frac{\sigma_x^2}{\sigma_n^2} S_h(f) + 1 \right) df
\end{aligned} \tag{127}$$

where for (a) we have used (32), and the fact that for constant modulus input symbols  $h'(\mathbf{y}|\mathbf{x})$  is equal to (45), see (46). Furthermore,  $\mathcal{I}(y; x|h) |_{\text{CM}, \sigma_x^2}$  corresponds to the coherent mutual information using circularly symmetric constant modulus input symbols with power  $\sigma_x^2$ .

Hence, we have found a lower bound on the capacity that is achievable with i.i.d. constant modulus input symbols with a uniformly distributed phase. However, as far as we know there is no closed form solution for the first term in (127), i.e.,  $\mathcal{I}(y; x|h) |_{\text{CM}, \sigma_x^2}$ , so it has to be calculated numerically. In addition, for nominal peak-to-average power ratios  $\beta > 1$  this bound is in general not tight. The lower bound (127) in combination with (84) is shown in Fig. 3. As it is based on constant modulus signaling, this bound becomes loose with an increasing SNR. The lower bound in (127) corresponds to the lower bound on the peak power constrained capacity given in [2, (34)].

Using the well known time-sharing argument, the lower bound on capacity given in (127) can be enhanced. The time-sharing argumentation is based on the fact that, while keeping the average transmit power constant, using the channel only during a fraction of the time might lead to a higher achievable rate. Using this time-sharing argument, a lower bound on the peak power constrained capacity for input

distributions with an average power  $\sigma_x^2$  and a nominal peak-to-average power ratio  $\beta$  is consequently given by the following expression:

$$\sup_{\mathcal{P}_{\text{peak}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) \geq \max_{\gamma \in [1, \beta]} \left\{ \frac{1}{\gamma} \mathcal{I}(y; x|h) |_{\text{CM}, \gamma \sigma_x^2} - \frac{1}{\gamma} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left( \frac{\gamma \sigma_x^2}{\sigma_n^2} S_h(f) + 1 \right) df \right\}. \quad (128)$$

This lower bound exactly corresponds to the lower bound on the peak power constrained capacity given in [2, (34)/(29)].<sup>8</sup> As the lower bound in (128) does not hold for i.i.d. input symbols due to the application of the time-sharing argument, it would be unfair to use it for comparison in Fig. 3. Thus, in Fig. 3 (127) is shown.

Note that in contrast to the lower bound on the achievable rate for i.i.d. zero-mean proper Gaussian input distributions in (82), the lower bound in (128) does not converge to the coherent capacity for asymptotically small channel dynamics, i.e.,  $f_d \rightarrow 0$ , as the coherent mutual information rate, which equals  $\mathcal{I}(y; x|h)$ , with any peak power limited input distribution is smaller than the coherent capacity, which is achieved for i.i.d. zero-mean proper Gaussian input symbols, cf. (31). This is also one advantage of our study of bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols. As the coherent capacity is achieved by this input distribution, this approach allows to give a lower bound on the achievable rate, which becomes tight for asymptotically small channel dynamics.

#### IV. ALTERNATIVE UPPER BOUND ON THE ACHIEVABLE RATE WITH I.I.D. INPUT SYMBOLS BASED ON THE ONE-STEP CHANNEL PREDICTION ERROR VARIANCE

In Section III, we have derived bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols and also have discussed their link to capacity bounds for peak power constrained input symbols given in [2] and [1]. These bounds are based on a purely mathematical derivation and do not give any link to a physical interpretation like the channel prediction error variance as it has been used in [3]. In the present section, we give a new upper bound on the achievable rate which is based on the channel prediction error variance and is also not restricted to peak power constrained input symbols. In contrast, for the derivation of the channel prediction based capacity bounds in [3], the peak power constraint has been required for technical reasons. However, for this derivation we have to restrict to i.i.d. input symbols, which has not been required for the derivation of the upper bound on the achievable rate in Section III<sup>9</sup>. As no peak power constraint is required for the derivation of the upper bound in the present section, we are able to evaluate the new upper bound also for i.i.d. zero-mean proper Gaussian input symbols. Additionally, we will also evaluate the upper bound for peak power constrained input symbols. However, due to the required restriction to i.i.d. input symbols, the resulting upper bound with peak power constrained input symbols is not an upper bound on the peak power constrained capacity, but only on the achievable rate with i.i.d. input symbols and a peak power constraint. In contrast to the upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given in Section III-D2, which holds only for a rectangular PSD of the fading process, the upper bound given now holds for an arbitrary PSD with compact support.

In the first part of the following derivation, we only restrict to i.i.d. input symbols with a maximum average power  $\sigma_x^2$ . Any other restriction on the input symbols, either to zero-mean proper Gaussian symbols or a peak power constraint will be applied later. Thus, we define the set of all i.i.d. input distributions

<sup>8</sup>Note that it would also be possible to enhance the lower bound on the capacity for zero-mean proper Gaussian inputs in (82) based on the time-sharing argument, i.e., by discarding the restriction to identically distributed input symbols. However, as for the derivation of the upper bound on the achievable rate in (85) we need the restriction to i.i.d. input symbols, such a lower bound without the assumption on i.i.d. input symbols would not match this upper bound. Therefore, we do not consider this further.

<sup>9</sup>Note, in Section III the upper bound has only been evaluated for i.i.d. zero-mean proper Gaussian input symbols in the final step, a restriction to independent input symbols is not required for the derivation itself.

with a maximum average power  $\sigma_x^2$  as

$$\mathcal{P}_{\text{i.i.d.}} = \left\{ p(\mathbf{x}) \middle| \mathbf{x} \in \mathbb{C}^N, p(\mathbf{x}) = \prod_{i=1}^N p(x_i), p(x_i) = p(x_j) \forall i, j, \mathbb{E}[|x_k|^2] \leq \sigma_x^2 \quad \forall k \right\}. \quad (129)$$

#### A. Achievable Rate based on Channel Prediction

Corresponding to Section III, we express the mutual information rate  $\mathcal{I}'(\mathbf{y}; \mathbf{x})$  based on the separation in (27). As previously stated, we construct an upper bound on the achievable rate based on channel prediction. As the channel fading process is stationary and ergodic, and as we assume i.i.d. input symbols, we can rewrite  $h'(\mathbf{y}|\mathbf{x})$  as follows:

$$\begin{aligned} h'(\mathbf{y}|\mathbf{x}) &= \lim_{N \rightarrow \infty} \frac{1}{N} h(\mathbf{y}|\mathbf{x}) \\ &\stackrel{(a)}{=} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N h(y_k | \mathbf{x}, \mathbf{y}_1^{k-1}) \\ &\stackrel{(b)}{=} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N h(y_k | \mathbf{x}_1^k, \mathbf{y}_1^{k-1}) \\ &\stackrel{(c)}{=} \lim_{N \rightarrow \infty} h(y_N | \mathbf{x}_1^N, \mathbf{y}_1^{N-1}) \end{aligned} \quad (130)$$

where, e.g., the vector  $\mathbf{y}_1^{N-1}$  contains all channel output symbols from the time instant 1 to the time instant  $N - 1$ . Here, for (a) we have used the chain rule for differential entropy, (b) is based on the fact that  $y_k$  conditioned on  $\mathbf{y}_1^{k-1}$  and  $\mathbf{x}_1^k$  is independent of the symbols  $\mathbf{x}_{k+1}^N$  due to the independency of the transmit symbols. Equality (c) follows from the ergodicity and stationarity of the channel fading process and the assumption on independent transmit symbols, see [28, Chapter 4.2]. Correspondingly,  $h'(\mathbf{y})$  can be rewritten as follows:

$$h'(\mathbf{y}) = \lim_{N \rightarrow \infty} h(y_N | \mathbf{y}_1^{N-1}). \quad (131)$$

Thus, based on (130) and (131), the achievable rate is given by

$$\mathcal{I}'(\mathbf{y}; \mathbf{x}) = \lim_{N \rightarrow \infty} \{h(y_N | \mathbf{y}_1^{N-1}) - h(y_N | \mathbf{x}_1^N, \mathbf{y}_1^{N-1})\} \quad (132)$$

which we name *prediction separation* of the mutual information rate.

#### B. An Upper Bound based on the Channel Prediction Error Variance

Now, we will upper-bound the achievable rate based on the expression in (132).

1) *Upper Bound on  $h'(\mathbf{y})$* : As conditioning reduces entropy, we can upper bound  $h(y_N | \mathbf{y}_1^{N-1})$  in (131) by

$$h(y_N | \mathbf{y}_1^{N-1}) \leq h(y_N). \quad (133)$$

Using (131), (133), ergodicity, and stationarity, we get

$$h'(\mathbf{y}) \leq h(y_N) \stackrel{(a)}{\leq} \log(\pi e (\alpha \sigma_x^2 \sigma_h^2 + \sigma_n^2)) = h'_U(\mathbf{y}) \quad (134)$$

where for (a) we used the fact that proper Gaussian distributions maximize entropy and that the average transmit power is given by  $\alpha \sigma_x^2$  with  $\alpha \in [0, 1]$ . Using an average transmit power of  $\alpha \sigma_x^2$  still enables to choose average transmit powers smaller than the maximum average transmit power  $\sigma_x^2$ .

Obviously, the upper bound on  $h'(\mathbf{y})$  in (134) is equal to the upper bound in (36), except of the factor  $\alpha$ , which we introduced here to account for average transmit powers smaller than the maximum average transmit power  $\sigma_x^2$ . This is relevant in case of peak power constrained input symbols, see Section III-G1.

2) *The Entropy Rate  $h'(\mathbf{y}|\mathbf{x})$* : In the following, we lower-bound  $h'(\mathbf{y}|\mathbf{x})$  based on the channel prediction representation in (130). This lower-bounding approach of  $h'(\mathbf{y}|\mathbf{x})$  is completely different to the one used in Section III-C2. Therefore, within the present section we express  $h(y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1})$  at the RHS of (130) based on the one-step channel prediction error variance. As the following argumentation will show, the channel output  $y_N$  conditioned on  $\mathbf{x}_1^N, \mathbf{y}_1^{N-1}$  is proper Gaussian and, thus, fully characterized by its conditional mean and conditional variance. The conditional mean is given by

$$\begin{aligned} \mathbb{E}[y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}] &= \mathbb{E}[x_N h_N + n_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}] \\ &= x_N \mathbb{E}[h_N|\mathbf{x}_1^{N-1}, \mathbf{y}_1^{N-1}] \\ &= x_N \hat{h}_N \end{aligned} \quad (135)$$

where  $\hat{h}_N$  is the MMSE estimate of  $h_N$  based on the channel output observations at all previous time instances and the channel input symbols at these time instances. Based on  $\hat{h}_N$ , the channel output  $y_N$  can be written as

$$y_N = x_N h_N + n_N = x_N (\hat{h}_N + e_N) + n_N \quad (136)$$

with the prediction error given by

$$e_N = h_N - \hat{h}_N. \quad (137)$$

As both, the noise as well as the fading process, are jointly proper Gaussian, the MMSE estimate is equivalent to the linear minimum mean squared error (LMMSE). Thus,  $\hat{h}_N$  and  $h_N$  are jointly proper Gaussian and it follows that the estimation error  $e_N$  is zero-mean proper Gaussian. Note that here  $\hat{h}_N$  also has zero mean.

As  $e_N$  is proper Gaussian, it can be easily seen by (136) that  $y_N$  conditioned on  $\mathbf{x}_1^N, \mathbf{y}_1^{N-1}$  is also proper Gaussian. Thus, for the evaluation of  $h(y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1})$ , we calculate the conditional variance of the channel output  $y_N$  which is given by

$$\begin{aligned} \text{var}[y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}] &= \mathbb{E}\left[|y_N - \mathbb{E}[y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}]|^2 \middle| \mathbf{x}_1^N, \mathbf{y}_1^{N-1}\right] \\ &= \mathbb{E}\left[|x_N(h_N - \hat{h}_N) + n_N|^2 \middle| \mathbf{x}_1^N, \mathbf{y}_1^{N-1}\right] \\ &= |x_N|^2 \mathbb{E}[|e_N|^2 \middle| \mathbf{x}_1^{N-1}, \mathbf{y}_1^{N-1}] + \sigma_n^2 \\ &= |x_N|^2 \sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1}) + \sigma_n^2 \end{aligned} \quad (138)$$

where

$$\begin{aligned} \sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1}) &= \mathbb{E}\left[|h_N - \hat{h}_N|^2 \middle| \mathbf{x}_1^{N-1}, \mathbf{y}_1^{N-1}\right] \\ &= \mathbb{E}\left[|e_N|^2 \middle| \mathbf{x}_1^{N-1}, \mathbf{y}_1^{N-1}\right] \\ &\stackrel{(a)}{=} \mathbb{E}\left[|e_N|^2 \middle| \mathbf{x}_1^{N-1}\right] \end{aligned} \quad (139)$$

is the prediction error variance of the MMSE estimator for  $\hat{h}_N$ . For (a) we have used the fact that the zero-mean estimation error  $e_N$  is orthogonal to and, thus, independent of the observations  $\mathbf{y}_1^{N-1}$ . However, the prediction error variance depends on the input symbols  $\mathbf{x}_1^{N-1}$ , which is indicated by writing  $\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1})$ .

Based on the channel prediction error variance, we can rewrite the entropy  $h(y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1})$  as

$$\begin{aligned} h(y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}) &= \mathbb{E}_{\mathbf{x}} \left[ \log(\pi e \text{ var}[y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}]) \right] \\ &= \mathbb{E}_{\mathbf{x}} \left[ \log \left( \pi e \left( \sigma_n^2 + \sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1}) |x_N|^2 \right) \right) \right]. \end{aligned} \quad (140)$$

With (130) and (140), we get for i.i.d. input symbols

$$h'(\mathbf{y}|\mathbf{x}) = \mathbb{E}_{x_k} \left[ \mathbb{E}_{\mathbf{x}_{-\infty}^{k-1}} \left[ \log \left( \pi e \left( \sigma_n^2 + \sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1}) |x_k|^2 \right) \right) \right] \right] \quad (141)$$

where  $\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$  is the prediction error variance in (139) for an infinite number of channel observations in the past, i.e.,

$$\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1}) = \lim_{N \rightarrow \infty} \sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1}) \quad (142)$$

which is indicated by writing  $\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$ . Note that we have switched the notation and now predict at the time instant  $k$  instead of predicting at the time instant  $N$ . This is possible, as the channel fading process is stationary, the input symbols are assumed to be i.i.d., and as we consider an infinitely long past.

3) *Upper Bound on the Achievable Rate:* With (132), (133), (134), and (140)/(141), we can give the following upper bound on the achievable rate with i.i.d. input symbols:

$$\mathcal{I}'(\mathbf{y}; \mathbf{x}) \leq \log(\alpha\rho + 1) - \mathbb{E}_{x_k} \left[ \mathbb{E}_{\mathbf{x}_{-\infty}^{k-1}} \left[ \log \left( 1 + \frac{\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})}{\sigma_n^2} |x_k|^2 \right) \right] \right] \quad (143)$$

where  $\rho$  is the nominal average SNR, see (11). Obviously, the upper bound in (143) still depends on the channel prediction error variance  $\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$  given in (142), which itself depends on the distribution of the input symbols in the past. Effectively  $\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$  is itself a random quantity. For infinite transmission lengths, i.e.,  $N \rightarrow \infty$ , its distribution is independent of the time instant  $k$ , as the channel fading process is stationary and as the transmit symbols are i.i.d..

4) *The Prediction Error Variance  $\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$ :* The prediction error variance  $\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$  in (142) depends on the distribution of the input symbols  $\mathbf{x}_{-\infty}^{k-1}$ . To construct an upper bound on the RHS of (143), we need to find a distribution of the transmit symbols in the past, i.e.,  $\mathbf{x}_{-\infty}^{k-1}$ , which leads to a distribution of  $\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$  that maximizes the RHS of (143). Therefore, we have to express the channel prediction error variance  $\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$  as a function of the transmit symbols in the past, i.e.,  $\mathbf{x}_{-\infty}^{k-1}$ . In a first step, we give such an expression for the case of a finite past time horizon, i.e., for  $\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1})$  in (139) which can be expressed by

$$\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1}) = \sigma_h^2 - \mathbf{r}_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}}^H \mathbf{R}_{\mathbf{y}_1^{N-1}|\mathbf{x}_1^{N-1}}^{-1} \mathbf{r}_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}} \quad (144)$$

where  $\mathbf{R}_{\mathbf{y}_1^{N-1}|\mathbf{x}_1^{N-1}}$  is the correlation matrix of the observations  $\mathbf{y}_1^{N-1}$  while the past transmit symbols  $\mathbf{x}_1^{N-1}$  are known, i.e.,

$$\begin{aligned} \mathbf{R}_{\mathbf{y}_1^{N-1}|\mathbf{x}_1^{N-1}} &= \mathbb{E} [\mathbf{y}_1^{N-1} (\mathbf{y}_1^{N-1})^H | \mathbf{x}_1^{N-1}] \\ &= \mathbf{X}_{N-1} \mathbf{R}_h \mathbf{X}_{N-1}^H + \sigma_n^2 \mathbf{I}_{N-1} \end{aligned} \quad (145)$$

with  $\mathbf{X}_{N-1}$  being a diagonal matrix containing the past transmit symbols such that  $\mathbf{X}_{N-1} = \text{diag}(\mathbf{x}_1^{N-1})$ . In addition,  $\mathbf{R}_h$  is the autocorrelation matrix of the channel fading process

$$\mathbf{R}_h = \mathbb{E} [\mathbf{h}_1^{N-1} (\mathbf{h}_1^{N-1})^H] \quad (146)$$

where  $\mathbf{h}_1^{N-1}$  contains the fading weights from time instant 1 to  $N-1$ . The cross correlation vector  $\mathbf{r}_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}}$  between the observation vector  $\mathbf{y}_1^{N-1}$  and the fading weight  $h_N$  while knowing the past transmit symbols  $\mathbf{x}_1^{N-1}$  is given by

$$\mathbf{r}_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}} = \mathbb{E} [\mathbf{y}_1^{N-1} h_N^* | \mathbf{x}_1^{N-1}] = \mathbf{X}_{N-1} \mathbf{r}_{h,\text{pred}} \quad (147)$$

with  $\mathbf{r}_{h,\text{pred}} = [r_h(-(N-1)) \dots r_h(-1)]^T$  where  $r_h(l)$  is the autocorrelation function as defined in (2).

Substituting (145) and (147) into (144) yields

$$\begin{aligned}
\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1}) &= \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \mathbf{X}_{N-1}^H (\mathbf{X}_{N-1} \mathbf{R}_h \mathbf{X}_{N-1}^H + \sigma_n^2 \mathbf{I}_{N-1})^{-1} \mathbf{X}_{N-1} \mathbf{r}_{h,\text{pred}} \\
&= \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \left( \mathbf{R}_h + \sigma_n^2 (\mathbf{X}_{N-1}^H \mathbf{X}_{N-1})^{-1} \right)^{-1} \mathbf{r}_{h,\text{pred}} \\
&\stackrel{(a)}{=} \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H (\mathbf{R}_h + \sigma_n^2 \mathbf{Z}^{-1})^{-1} \mathbf{r}_{h,\text{pred}}
\end{aligned} \tag{148}$$

where for (a) we have used  $\mathbf{Z} = \mathbf{X}_{N-1}^H \mathbf{X}_{N-1}$ , i.e.,  $\mathbf{Z}$  is a diagonal matrix containing the powers of the individual transmit symbols in the past from time instant 1 to  $N-1$ . For ease of notation, we omit the index  $N-1$ .<sup>10</sup>

Remember that we want to derive an upper bound on the achievable rate with i.i.d. input symbols by maximizing the RHS of (143) over all i.i.d. distributions of the transmit symbols in the past with an average power  $\alpha\sigma_x^2$ . Obviously, the distribution of the phases of the past transmit symbols  $\mathbf{x}_1^{N-1}$  has no influence on the channel prediction error variance  $\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1})$ . Thus, it rests to evaluate, for which distribution of the power of the past transmit symbols the RHS of (143) is maximized. In the following, we will show that the RHS of (143) is maximized in case the past transmit symbols have a constant power  $\alpha\sigma_x^2$ . I.e., calculation of the prediction error variance under the assumption that the past transmit symbols are constant modulus symbols with transmit power  $|x_k|^2 = \alpha\sigma_x^2$  maximizes the RHS of (143) over all i.i.d. input distributions for the given average power constraint of  $\alpha\sigma_x^2$ .

To prove this statement, we use the fact that the expression in the expectation operation at the RHS of (143) (but here for the case of a finite past time horizon) with (148), i.e.,

$$\log \left( 1 + \frac{|x_N|^2}{\sigma_n^2} \left( \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H (\mathbf{R}_h + \sigma_n^2 \mathbf{Z}^{-1})^{-1} \mathbf{r}_{h,\text{pred}} \right) \right) \tag{149}$$

is convex with respect to each individual element of the diagonal of  $\mathbf{Z}$ , which we name  $\mathbf{z}$ . The proof of convexity of (149) is given in Appendix D. Based on this convexity and Jensen's inequality, we get

$$\begin{aligned}
&\mathbb{E}_{\mathbf{z}} \left[ \log \left( 1 + \frac{|x_N|^2}{\sigma_n^2} \left( \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H (\mathbf{R}_h + \sigma_n^2 \mathbf{Z}^{-1})^{-1} \mathbf{r}_{h,\text{pred}} \right) \right) \right] \\
&\geq \log \left( 1 + \frac{|x_N|^2}{\sigma_n^2} \left( \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H (\mathbf{R}_h + \sigma_n^2 (\mathbb{E}_{\mathbf{z}} [\mathbf{Z}])^{-1})^{-1} \mathbf{r}_{h,\text{pred}} \right) \right) \\
&= \log \left( 1 + \frac{|x_N|^2}{\sigma_n^2} \left( \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \left( \mathbf{R}_h + \frac{\sigma_n^2}{\alpha\sigma_x^2} \mathbf{I}_{N-1} \right)^{-1} \mathbf{r}_{h,\text{pred}} \right) \right) \\
&= \log \left( 1 + \frac{|x_N|^2}{\sigma_n^2} \sigma_{e_{\text{pred,CM}}}^2 \right)
\end{aligned} \tag{150}$$

where  $\sigma_{e_{\text{pred,CM}}}^2$  is the channel prediction error variance in case all past transmit symbols are constant modulus symbols with power  $\alpha\sigma_x^2$ . Here, the index CM denotes constant modulus.

As this lower-bounding of (149) can be performed for an arbitrary  $N$ , i.e., for an arbitrarily long past, we can also conclude that the RHS of (143) is upper-bounded by

$$\mathcal{I}'(\mathbf{y}; \mathbf{x}) \leq \log(\alpha\rho + 1) - \mathbb{E}_{x_k} \left[ \log \left( 1 + \frac{\sigma_{e_{\text{pred,CM},\infty}}^2}{\sigma_n^2} |x_k|^2 \right) \right] \tag{151}$$

where  $\sigma_{e_{\text{pred,CM},\infty}}^2$  is the channel prediction error variance in case all past transmit symbols are constant modulus symbols with power  $\alpha\sigma_x^2$  and an infinitely long past observation horizon. In this case, the prediction error variance is no longer a random quantity but is constant for all time instances  $k$ .

<sup>10</sup>Note that the inverse of  $\mathbf{Z}$  in (148) does not exist, if a diagonal element  $z_i$  of the diagonal matrix  $\mathbf{Z}$  is zero, i.e., one transmit symbol has zero power. However, as the prediction error variance is continuous in  $z_i = 0$  for all  $i$  this does not lead to problems in the following derivation.

Constant modulus symbols are in general not the capacity maximizing input distribution. However, we only use them to find a distribution of  $\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$  that maximizes (143).

For constant modulus input symbols with power  $\alpha\sigma_x^2$  and an infinitely long past, the prediction error variance is given by, cf. [3]

$$\sigma_{e_{\text{pred},\text{CM},\infty}}^2 = \frac{\sigma_n^2}{\alpha\sigma_x^2} \left\{ \exp \left( \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left( 1 + \frac{\alpha\sigma_x^2}{\sigma_n^2} S_h(f) \right) df \right) - 1 \right\} \quad (152)$$

As far as we know, the upper bound on the achievable rate in (151) is new. The innovation in the derivation of this bound lies in the fact that we separate the input symbols into the one at the time instant  $x_k$  and the previous input symbols contained in  $\mathbf{x}_{-\infty}^{k-1}$ . The latter ones are only relevant to calculate the prediction error variance, which itself is a random variable depending on the distribution of the past transmit symbols. To derive an upper bound on the achievable rate with i.i.d. input distributions, we have shown that the achievable rate is upper-bounded if the prediction error variance is calculated under the assumption that all past transmit symbols are constant modulus input symbols. As the assumption on constant modulus symbols is only used in the context of the prediction error variance, the upper bound on the achievable rate still holds for any i.i.d. input distribution with the given average power constraint. This allows us to evaluate this bound also for the case of i.i.d. zero-mean proper Gaussian input symbols.

5) *Proper Gaussian Input Symbols:* Evaluating (151) for i.i.d. zero-mean proper Gaussian (PG) input symbols yields

$$\begin{aligned} \mathcal{I}'(\mathbf{y}; \mathbf{x})|_{\text{PG}} &\leq \sup_{\alpha \in [0,1]} \left\{ \log(\alpha\rho + 1) - \int_0^\infty \log \left( 1 + \frac{\sigma_{e_{\text{pred},\text{CM},\infty}}^2}{\sigma_h^2} \alpha\rho z \right) e^{-z} dz \right\} \\ &\stackrel{(a)}{\leq} \sup_{\alpha \in [0,1]} \left\{ \log(\alpha\rho + 1) - \int_0^\infty \log \left( 1 + \frac{\sigma_{e_{\text{pred},\text{CM},\infty}}^2|_{\alpha=1}}{\sigma_h^2} \alpha\rho z \right) e^{-z} dz \right\} \\ &\stackrel{(b)}{=} \log(\rho + 1) - \int_0^\infty \log \left( 1 + \frac{\sigma_{e_{\text{pred},\text{CM},\infty}}^2|_{\alpha=1}}{\sigma_h^2} \rho z \right) e^{-z} dz \\ &= \mathcal{I}'_U(\mathbf{y}; \mathbf{x})|_{\text{pred,PG}} \end{aligned} \quad (153)$$

where (a) is based on the fact that  $\sigma_{e_{\text{pred},\text{CM},\infty}}^2$  monotonically decreases with an increasing  $\alpha$ , and, thus, that the term in the second line of (153) is maximized if the prediction error variance is calculated for  $\alpha = 1$ , which is denoted by writing  $\sigma_{e_{\text{pred},\text{CM},\infty}}^2|_{\alpha=1}$ . Furthermore, (b) follows from the monotonicity of the argument of the supremum in the third line of (153) in  $\alpha$ , which can be shown analogously to the monotonicity of (88) based on (90). In conclusion, this means that the upper bound for zero-mean proper Gaussian input symbols is maximized for the maximum average transmit power  $\sigma_x^2$ .

As the coherent mutual information rate  $\mathcal{I}'(\mathbf{y}; \mathbf{x}|\mathbf{h})$  upper-bounds  $\mathcal{I}'(\mathbf{y}; \mathbf{x})$ , we can enhance the upper bound in (153) analogously to (87) with  $\mathcal{I}'(\mathbf{x}; \mathbf{y}|\mathbf{h})$  given in (31).

Fig. 4 shows the prediction based upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given in (153) in comparison to the upper and lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs given Section III-D for a rectangular PSD of the channel fading process. Both upper bounds are shown in combination with the coherent upper bound, i.e., (87) and (31). A comparison of the prediction based upper bound (153)/(87) and the bound given in (85)/(87) shows, that it depends on the channel parameters which one is tighter. It can easily be shown that for  $f_d \rightarrow 0$  and for  $f_d = 0.5$  both bounds, i.e., (85) and (153), are equal. For other  $f_d$  it depends on the SNR  $\rho$  which bound is tighter. An analytical comparison turns out to be difficult as in both cases we use a different way of lower-bounding  $h'(\mathbf{y}|\mathbf{x})$ .



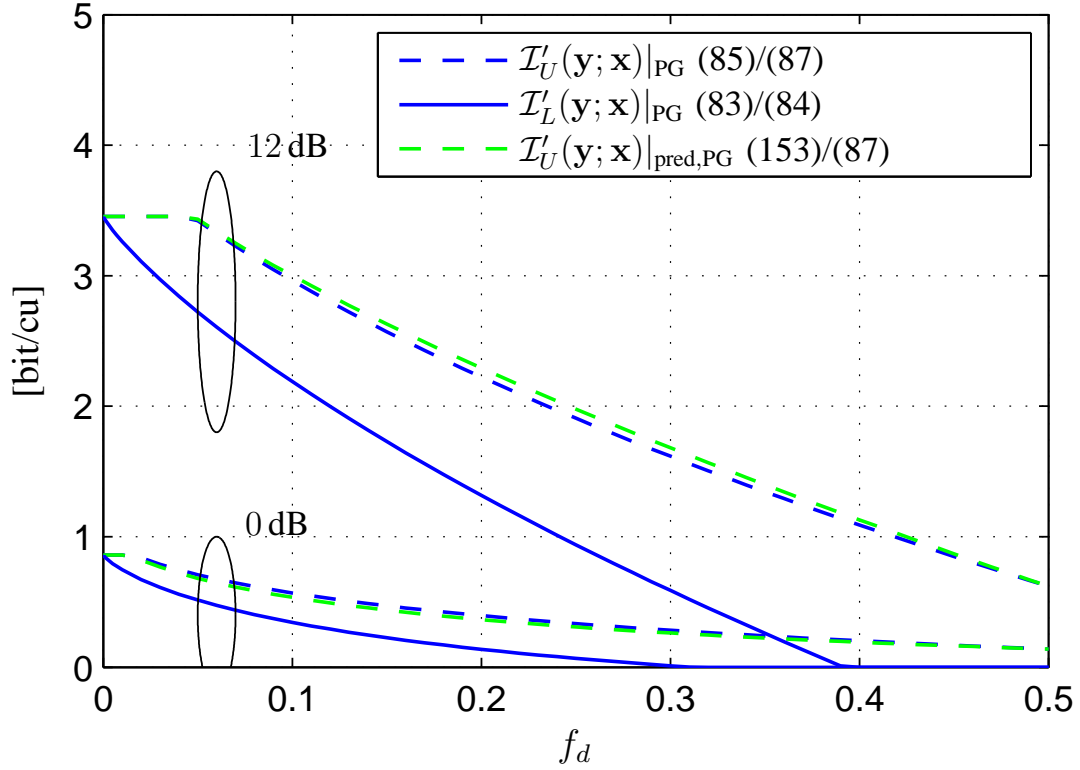


Fig. 4. Comparison of the upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs based on channel prediction (153)/(87) with the upper bound given in (85)/(87); in addition the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs (83)/(84) is shown; rectangular PSD  $S_h(f)$

6) *Peak Power Constrained Input Distributions:* Now, we consider the case of a peak power constrained to  $P_{\text{peak}}$  in addition to the average power constraint. With the nominal peak-to-average power ratio  $\beta = P_{\text{peak}}/\sigma_x^2$ , with (151) we get the following upper bound on the achievable rate with i.i.d. input symbols:

$$\begin{aligned} \sup_{\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) &\leq \sup_{\alpha \in [0,1]} \sup_{\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}|_{\alpha}} \left\{ \log(\alpha\rho + 1) - \mathbb{E}_{x_k} \left[ \log \left( 1 + \frac{\sigma_{e_{\text{pred,CM},\infty}}^2}{\sigma_n^2} |x_k|^2 \right) \right] \right\} \\ &\stackrel{(a)}{\leq} \sup_{\alpha \in [0,1]} \left\{ \log(\alpha\rho + 1) - \frac{\alpha}{\beta} \log \left( 1 + \frac{\sigma_{e_{\text{pred,CM},\infty}}^2}{\sigma_h^2} \rho\beta \right) \right\} \end{aligned} \quad (154)$$

where  $\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}$  corresponds to  $\mathcal{P}_{\text{i.i.d.}}$  but with the additional peak power constraint  $|x_k|^2 \leq \beta\sigma_x^2$ .  $\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}|_{\alpha}$  corresponds to  $\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}$  in (129) but with the average transmit power fixed to  $\alpha\sigma_x^2$ . Inequality (a) can be shown following an analogous argumentation as in Section III-G1 from (113) to (117). Note that the prediction error variance  $\sigma_{e_{\text{pred,CM},\infty}}^2$  depends on  $\alpha$ . Now, we would have to calculate the supremum of the RHS of (154) with respect to  $\alpha$  which turns out to be difficult due to the dependency of  $\sigma_{e_{\text{pred,CM},\infty}}^2$  on  $\alpha$ . However, as  $\sigma_{e_{\text{pred,CM},\infty}}^2$  monotonically decreases with an increasing  $\alpha$ , and as the RHS of (154) monotonically increases with a decreasing  $\sigma_{e_{\text{pred,CM},\infty}}^2$ , we can upper-bound (154) by setting  $\alpha = 1$  within  $\sigma_{e_{\text{pred,CM},\infty}}^2$  in (152), i.e.,

$$\begin{aligned} \sup_{\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}} \mathcal{I}'(\mathbf{y}; \mathbf{x}) &\leq \sup_{\alpha \in [0,1]} \left\{ \log(\alpha\rho + 1) - \frac{\alpha}{\beta} \log \left( 1 + \frac{\sigma_{e_{\text{pred,CM},\infty}}^2|_{\alpha=1}}{\sigma_h^2} \rho\beta \right) \right\} \\ &= \log(\alpha_{\text{opt}}\rho + 1) - \frac{\alpha_{\text{opt}}}{\beta} \log \left( 1 + \frac{\sigma_{e_{\text{pred,CM},\infty}}^2|_{\alpha=1}}{\sigma_h^2} \rho\beta \right) \end{aligned}$$

$$= \mathcal{I}'_U(\mathbf{y}; \mathbf{x}) \big|_{\text{pred}, P_{\text{peak}}} \quad (155)$$

with

$$\alpha_{\text{opt}} = \min \left\{ 1, \left( \frac{1}{\beta} \log \left( 1 + \frac{\sigma_{e_{\text{pred}, \text{CM}, \infty}}^2 |_{\alpha=1}}{\sigma_h^2} \rho \beta \right) \right)^{-1} - \frac{1}{\rho} \right\}. \quad (156)$$

For (156), we have used the fact that the argument of the supremum in the first line of (155) is concave in  $\alpha$ , and, thus, there exists a unique maximum.

*a) Comparison to Capacity Bounds in [1] and [2]:* In the following, we will compare the upper bound on the achievable rate with peak power constrained i.i.d. input symbols in (155) with the upper bound on the peak power constrained capacity given in [1, Prop. 2.2], which is, modified to our notation, given by

$$C \leq \log(\alpha_{\text{opt}} \rho + 1) - \frac{\alpha_{\text{opt}}}{\beta} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left( \rho \beta \frac{S_h(f)}{\sigma_h^2} + 1 \right) df \quad (157)$$

with

$$\alpha_{\text{opt}} = \min \left\{ 1, \left( \frac{1}{\beta} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left( \rho \beta \frac{S_h(f)}{\sigma_h^2} + 1 \right) df \right)^{-1} - \frac{1}{\rho} \right\}. \quad (158)$$

Note that, in terms of the analytical expression, (157) corresponds to (120) for the special case of a rectangular PSD  $S_h(f)$ , see the discussion in Section III-G1.

On the other hand, we compare the upper bound on the achievable rate with peak power constrained i.i.d. input symbols in (155) with the lower bound on the peak power constrained capacity given in [2, (35)]

$$C_{l1}(\rho) = h(y_k | \hat{h}_k) - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left( \pi e \sigma_n^2 \left( 1 + \rho \frac{S_h(f)}{\sigma_h^2} \right) \right) df \quad (159)$$

where  $k$  is an arbitrary chosen time instant with an infinitely long past and  $h(y_k | \hat{h}_k)$  is the differential output entropy while conditioning on the channel estimate  $\hat{h}_k$ , being given by the MMSE estimate  $\mathbb{E}[h_k | \mathbf{x}_{-\infty}^{k-1}, \mathbf{y}_{-\infty}^{k-1}]$ , which is linear due to the fact that this estimation problem is jointly proper Gaussian. Based on the time-sharing argumentation, see Section III-G2, an enhanced lower bound on the peak power constrained capacity is given by [2, (29)/(35)]

$$C \geq \max_{\gamma \in [1, \beta]} \frac{1}{\gamma} C_{l1}(\rho \gamma). \quad (160)$$

*b) Numerical Evaluation:* Fig. 5 shows the upper bound on the achievable rate with i.i.d. input symbols and a peak power constraint based on the channel prediction error variance in (155)/(87) in comparison to the upper bound on the peak power constrained capacity given in [1, Prop. 2.2], i.e., (157), combined with (87), with  $\beta = 2$  for both. For comparison we use the lower bound on the peak power constrained capacity given in [2, (35)], i.e., (159) based on a constant modulus input distribution with 100 discrete signaling points with a uniform angular spacing. This approximates the case of a uniformly distributed phase. This lower bound is shown without time-sharing and with time-sharing ( $\gamma_{\text{opt}}$ ), see (160)<sup>11</sup>. Note that the lower bound in (159) is achievable with constant modulus input symbols with a uniformly distributed phase. Recall that time-sharing means, that the transmitter uses the channel only a  $1/\gamma$  part of the time. Obviously, time-sharing is not in accordance with the assumption on i.i.d. input symbols.

<sup>11</sup>Concerning the relation of the lower bound on the peak power constrained capacity in [2, (35)/(29)], i.e., (160)/(159), and the one used for comparison in Section III-G2, i.e., (127) respectively (128), see [2].

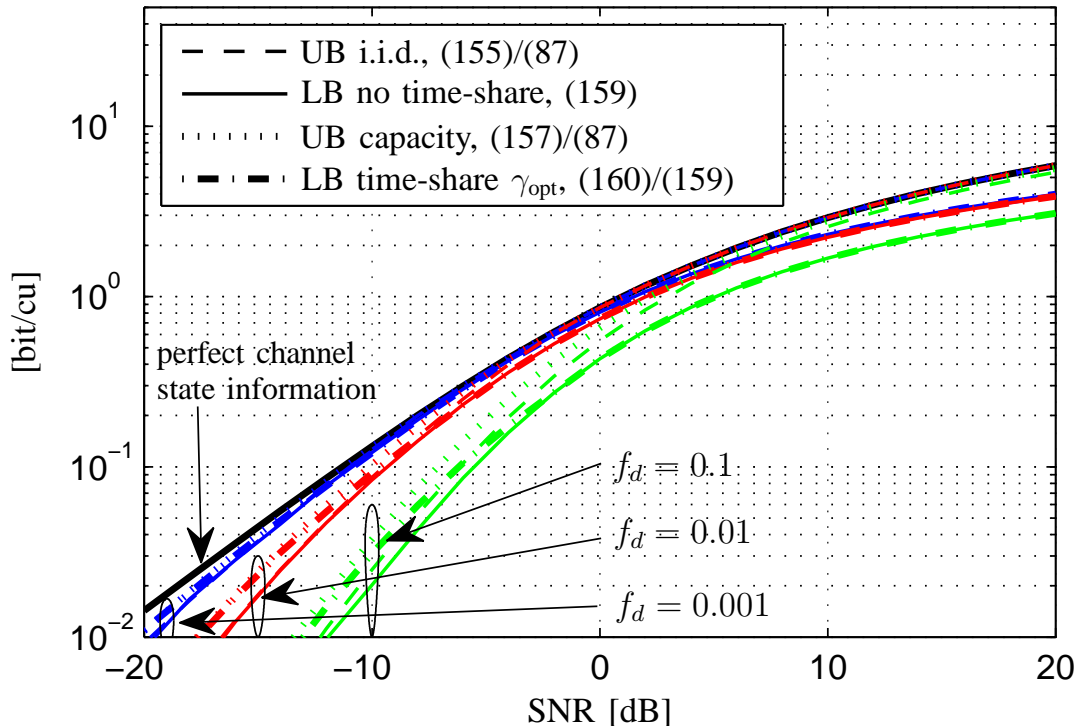


Fig. 5. Comparison of the upper bound on the achievable rate with i.i.d. symbols and a peak power constraint given in (155)/(87) based on channel prediction to the upper bound on the peak power constrained capacity given in [1, Proposition 2.2], i.e., (157)/(87), for  $\beta = 2$ ; in addition, the lower bound on the peak power constrained capacity [2, (35)], i.e., (159), is shown for a constant modulus (CM) input distribution with 100 signaling points, without and with time-sharing (160) ( $\gamma_{\text{opt}}$ ); rectangular PSD  $S_h(f)$

Therefore, the lower bound without time-sharing matches the new upper bound on the achievable rate with i.i.d. input symbols in (155)/(87), while the lower bound with time-sharing ( $\gamma_{\text{opt}}$ ) only matches the capacity upper bound in [1, Prop. 2.2], i.e., (157)/(87). From Fig. 5 it can be seen that the upper bound on the achievable rate with i.i.d. input symbols in (155)/(87) is lower or equal than the capacity upper bound in [1, Prop. 2.2], i.e., (157)/(87). However, (155)/(87) is only an upper bound on the achievable rate with i.i.d. input symbols and not an upper bound on the capacity, as i.i.d. input symbols are in general not capacity achieving, see [1] and Section III-G1. This can also be seen, as the lower bound on the achievable rate with time-sharing is larger than the upper bound on the achievable rate with i.i.d. input symbols (155)/(87) for very low SNRs. Furthermore, it is worth mentioning that for the case of a nominal peak-to-average power ratio  $\beta = 1$ , the upper bound in (155) and the one given in [1, Prop. 2.2], i.e., (157), coincide. In addition, the prediction based upper bound on the achievable rate in (155) as well as the capacity upper bound in [1, Prop. 2.2], i.e., (157), both become loose for  $\beta > 1$  and high SNR or  $\beta$  very large.

## V. COMPARISON TO SYNCHRONIZED DETECTION WITH A PILOT BASED CHANNEL ESTIMATION

In typical mobile communication systems periodical pilot symbols are introduced into the transmit data sequence. The pilot symbol spacing  $L$  is chosen such that the channel fading process is sampled at least with Nyquist frequency, i.e.,  $L < \lfloor 1/(2f_d) \rfloor$ . Based on these pilot symbols the channel is estimated, allowing for a coherent detection (synchronized detection). In conventional receivers, the channel estimation and the detection/decoding are two separate steps, such that the channel is estimated solely based on the pilot symbols. The resulting channel estimation error process is temporally correlated. However, performing coherent detection, the information contained in this temporal correlation is discarded. For a detailed discussion on this, we refer to [29]. The channel estimation error leads to an SNR degradation.

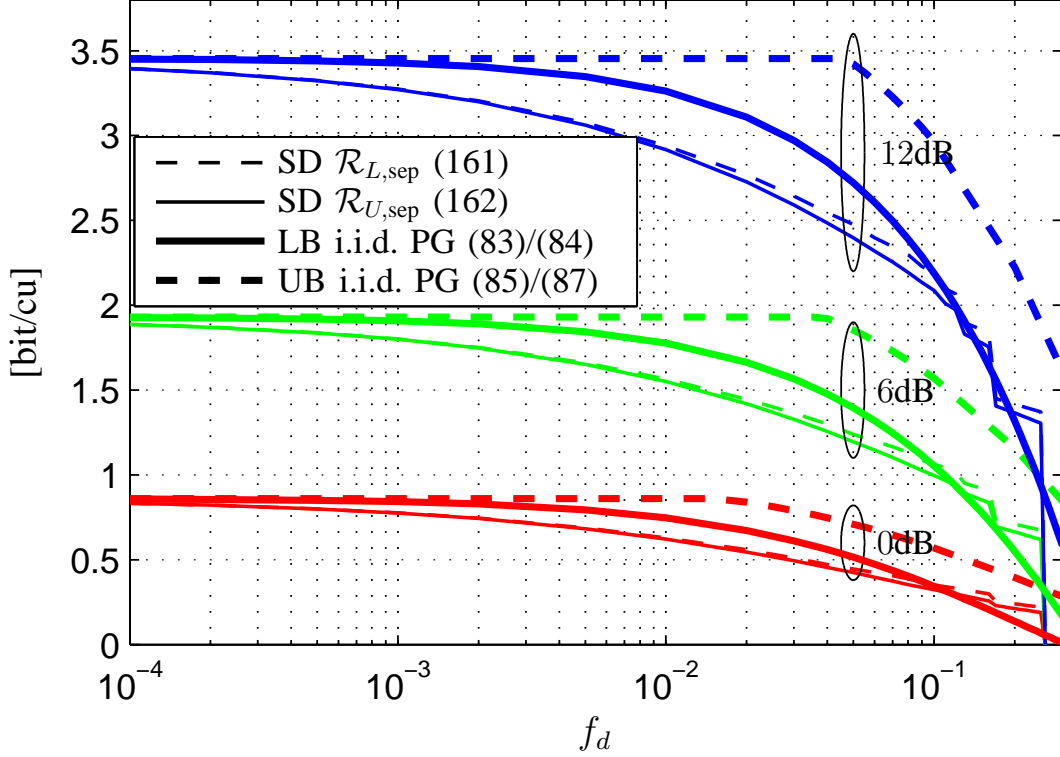


Fig. 6. Comparison of the bounds on the achievable rate with synchronized detection and a solely pilot based channel estimate (SD) and the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian (PG) input symbols given in (85)/(87) and (83)/(84); the pilot spacing  $L$  for synchronized detection is chosen such that  $\mathcal{R}_{L,\text{sep}}$  is maximized; rectangular PSD  $S_h(f)$

Bounds on the achievable rate for this separate processing have been given in [5]. For i.i.d. zero-mean proper Gaussian data symbols these bounds become

$$\mathcal{R}_{\text{sep}} \geq \mathcal{R}_{L,\text{sep}} = \frac{L-1}{L} \int_{z=0}^{\infty} \log \left( 1 + \rho \frac{1 - \frac{\sigma_{e,\text{pil}}^2}{\sigma_h^2}}{1 + \rho \frac{\sigma_{e,\text{pil}}^2}{\sigma_h^2}} z \right) e^{-z} dz \quad (161)$$

$$\mathcal{R}_{\text{sep}} \leq \mathcal{R}_{U,\text{sep}} = \mathcal{R}_{L,\text{sep}} + \frac{L-1}{L} \left( \log \left( 1 + \rho \frac{\sigma_{e,\text{pil}}^2}{\sigma_h^2} \right) - \int_{z=0}^{\infty} \log \left( 1 + \rho \frac{\sigma_{e,\text{pil}}^2}{\sigma_h^2} z \right) e^{-z} dz \right). \quad (162)$$

where  $\sigma_{e,\text{pil}}^2$  is the channel estimation error variance when estimating the channel solely based on pilot symbols which is given by

$$\sigma_{e,\text{pil}}^2 = \int_{f=-\frac{1}{2}}^{\frac{1}{2}} \frac{S_h(f)}{\frac{\rho}{L} \frac{S_h(f)}{\sigma_h^2} + 1} df. \quad (163)$$

Based on the lower bound in (161) it can easily be seen that the achievable rate is decreased in comparison to perfect channel knowledge by two factors. First, symbol time instances that are used for pilot symbols are lost for data symbols leading to the pre-log factor  $\frac{L-1}{L}$ , and secondly, the average SNR is decreased by the factor  $\left( 1 - \frac{\sigma_{e,\text{pil}}^2}{\sigma_h^2} \right) / \left( 1 + \rho \frac{\sigma_{e,\text{pil}}^2}{\sigma_h^2} \right)$  due to the channel estimation error variance.

Fig. 6 shows a comparison of the bounds on the achievable rate with synchronized detection based on a solely pilot based channel estimate in (161) and (162) with the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given in Section III-D. For synchronized detection with a solely

pilot based channel estimation the pilot spacing has been chosen such that the lower bound on the achievable rate in (161) is maximized. As this lower bound is relatively tight, the chosen pilot spacing should be close to the one that maximizes the achievable rate with synchronized detection using a solely pilot based channel estimation. Obviously, for the practical important range of small channel dynamics, i.e.,  $f_d \ll 0.1$ , the achievable rate with synchronized detection using a solely pilot based channel estimation stays below the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, indicating the possible gain when using enhanced receiver structures. Even in case of using pilot symbols, the receiver performance can be enhanced by using a joint processing of pilot and data symbols instead of a separate processing. For a more detailed discussion on the difference between separate and joint processing we refer to [29]. In this work also a lower bound on the achievable rate with joint processing of pilot and data symbols is given. One possibility of such a joint processing is to use an iterative code-aided channel estimation, where the channel estimation is enhanced based on reliability information on the data symbols delivered by the decoder. Based on this enhanced channel estimation detection and decoding is performed again, see e.g., [17] and [30].

## VI. CONCLUSION

The main focus of the present paper is the study of the achievable rate with i.i.d. zero-mean proper Gaussian input symbols on stationary Rayleigh flat-fading channels, where it is assumed that the receiver is aware of the law of the channel, but does not know its realization. We are interested in the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, as this input distribution serves well to upper-bound the achievable rate with practical modulation and coding schemes.

In the first part of this paper, i.e., in Section III, we have given a new upper bound on the achievable rate for i.i.d. zero-mean proper Gaussian input symbols, which holds in case of a rectangular PSD of the channel fading process. Furthermore, we also give a lower bound on the capacity which is achievable with i.i.d. zero-mean proper Gaussian input symbols. This lower bound is already known from [16]. With the upper and lower bound on the achievable rate for i.i.d. zero-mean proper Gaussian inputs, we have found a set of bounds, which is tight in the sense that their difference is bounded. We are able to bound this gap analytically by  $(1 + 2f_d)\gamma$  with the Euler constant  $\gamma \approx 0.577[\text{nat/cu}]$ . Thus, for the specific case of proper Gaussian inputs we give bounds, which are tight (in the sense given above) over the whole SNR range. In contrast, available bounds on capacity often focus only on a specific SNR range, e.g., [1] discusses the low SNR regime whereas [3] considers the high SNR regime.

The main novelty in this part of the paper lies in the new upper bound. It is based on a new lower bound on the conditional channel output entropy rate  $h'(\mathbf{y}|\mathbf{x})$  for the special case of a rectangular PSD of the channel fading process. This bound is not based on a peak power constraint, and, therefore, allows to give an upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs. To the best of our knowledge, this is the only known upper bound on the achievable rate without a peak power constraint, which is tight in the sense that its slope (pre-log) corresponds to the slope of the lower bound on the capacity. However, for the derivation of our upper bound on the achievable rate we need the restriction to a rectangular PSD of the channel fading process.

Furthermore, the comparison of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols with the asymptotic bounds on the peak power constrained capacity given in [3] shows the interesting fact that the achievable rate with i.i.d. zero-mean proper Gaussian inputs is characterized by the same asymptotic high SNR slope as the peak power constrained capacity. This shows that this kind of input distribution is not highly suboptimal with respect to its high SNR performance.

Moreover, we have discussed the relation of the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols to known bounds on the capacity with peak power constrained input symbols given in [2] and [1]. With respect to this, based on the given lower bound on  $h'(\mathbf{y}|\mathbf{x})$ , we have also derived an upper bound on the achievable rate with identically distributed (i.d.) peak power constrained input symbols, which is identified to be similar to an upper bound on capacity given in [1]. The assumption on i.d. input symbols is required in the derivation of our lower bound on  $h'(\mathbf{y}|\mathbf{x})$ . However,

due to this restriction, with our derivation we are not able to show that the given upper bound on the achievable rate is an upper bound on the peak power constrained capacity. Furthermore, our derivation is restricted to a rectangular PSD of the channel fading process, whereas the upper bound on capacity given in [1] holds for an arbitrary PSD of the channel fading process. Concerning the lower bounds, the difference of the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols and the lower bound on the peak power constrained capacity given in [2] results mainly from the coherent mutual information, which is part of the lower bound in both cases.

In the second part of the present paper, i.e., in Section IV, we have derived an alternative upper bound on the achievable rate with i.i.d. input symbols based on a prediction separation of the mutual information rate. Based on this separation, the conditional channel output entropy rate  $h'(y|x)$  can be expressed by the one-step channel prediction error variance, which is a well known result, see, e.g., [3]. We show for i.i.d. input symbols that the calculation of the prediction error variance  $\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$  under the assumption of constant modulus symbols yields an upper bound on the achievable rate. As the constant modulus assumption is only used in the context of  $\sigma_{e_{\text{pred},\infty}}^2(\mathbf{x}_{-\infty}^{k-1})$ , we can still give upper bounds on the achievable rate for general i.i.d. input symbol distributions, even for the case without a peak power constraint. On the one hand, we evaluate this upper bound for i.i.d. zero-mean proper Gaussian input symbols. It depends on the channel parameters if this upper bound based on the channel prediction error variance given in (153) or the upper bound derived in Section III-D2 is tighter. However, the prediction based bound is more general as it holds for arbitrary PSDs of the fading process with compact support and is not limited to rectangular PSDs as the one given in Section III-D2. On the other hand, we have evaluated the upper bound on the achievable rate based on the prediction error variance for peak power constrained input symbols. In this regard, we have observed that for nominal peak-to-average power ratios of  $\beta = 2$  and  $\beta = 1$  this upper bound on the achievable rate with i.i.d. input symbols is lower than or equal to the capacity upper bound in [1, Prop. 2.2]. But, it is not an upper bound on the capacity due to the restriction to i.i.d. input symbols. We do not know if this ordering holds in general.

Finally, in Section V, we have compared the bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols to bounds on the achievable rate with synchronized detection and a solely pilot based channel estimation. This comparison gives an indication of the possible gain when using enhanced receivers, e.g., receivers based on iterative code-aided channel estimation.

## APPENDIX A

### APPROXIMATION OF A RECTANGULAR PSD BY AN ABSOLUTELY SUMMABLE AUTOCORRELATION FUNCTION

In this appendix, we show that the rectangular PSD in (7), whose autocorrelation function is not absolutely summable but only square summable, see (5), can be arbitrarily closely approximated by a PSD with an absolutely summable autocorrelation function, see (4).

The discrete-time autocorrelation function  $r_h(l)$  corresponding to the rectangular PSD  $S_h(f)|_{\text{Rect}}$  in (7), which is given by

$$r_h(l) = \sigma_h^2 \text{sinc}(2f_d l) \quad (164)$$

is not absolutely summable. However, the rectangular PSD can be arbitrarily closely approximated by a PSD with a shape corresponding to the transfer function of a raised cosine filter, i.e.,

$$S_h(f)|_{\text{RC}} = \begin{cases} \frac{\sigma_h^2}{2f_d} & |f| \leq (1 - \beta_{\text{ro}})f_d \\ \frac{\sigma_h^2}{4f_d} \left[ 1 - \sin\left(\frac{\pi(f-f_d)}{2\beta_{\text{ro}}f_d}\right) \right] & (1 - \beta_{\text{ro}})f_d < f \leq (1 + \beta_{\text{ro}})f_d \\ 0 & f_d(1 + \beta_{\text{ro}}) < |f| \leq 0.5 \end{cases} \quad (165)$$

which for  $\beta_{\text{ro}} \rightarrow 0$  approaches the rectangular PSD  $S_h(f)|_{\text{Rect}}$ . Furthermore, the discrete-time autocorrelation function corresponding to  $S_h(f)|_{\text{RC}}$  is given by

$$r_h(l) = \sigma_h^2 \text{sinc}(2f_d l) \frac{\cos(\beta_{\text{ro}} \pi 2f_d l)}{1 - 4\beta_{\text{ro}}^2 4f_d^2 l^2} \quad (166)$$

which for  $\beta_{\text{ro}} > 0$  is absolutely summable. Thus, the rectangular PSD in (7) can be arbitrarily closely approximated by a PSD with an absolutely summable autocorrelation function.

## APPENDIX B

### MODIFIED UPPER BOUND ON $h'(\mathbf{y})$ FOR GAUSSIAN INPUTS

In this appendix, we derive an alternative upper bound on the channel output entropy rate  $h'(\mathbf{y})$  for the case of i.i.d. zero-mean proper Gaussian input symbols, which is tighter than the one given in (36). This derivation is based on work given in [21], [31]. As its evaluation requires more complex numerical methods, we do not further use this bound, but give it for completeness of presentation.

Obviously, an upper bound on the entropy rate  $h'(\mathbf{y})$  is given by assuming an uncorrelated channel fading process, i.e., its correlation matrix is assumed to be diagonal. This can be easily shown based on the chain rule for differential entropy

$$\begin{aligned} h'(\mathbf{y}) &= \lim_{N \rightarrow \infty} \frac{1}{N} h(\mathbf{y}) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^{\infty} h(y_k | \mathbf{y}_1^{k-1}) \\ &\stackrel{(a)}{\leq} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^{\infty} h(y_k) \\ &\stackrel{(b)}{=} h(y_k) \end{aligned} \quad (167)$$

where for (a) we have used the fact that conditioning reduces entropy and (b) is due to the ergodicity and the stationarity of the channel fading process and the assumption on i.i.d. input symbols. The major difference between this upper bound and the upper bound given in (36) is that the latter one implicitly corresponds to the case that the channel observations  $y_k$  are proper Gaussian, while the RHS of (167) still corresponds to the actual channel output entropy of the individual time instances. The upper bounding in (167) only discards the temporal dependencies between the different observations.

In the following, we calculate the entropy  $h(y_k)$  for the case of zero-mean proper Gaussian input symbols with an average power  $\sigma_x^2$

$$\begin{aligned} h(y_k) &= -\mathbb{E}_{y_k} [\log(p(y_k))] \\ &= -\int_{\mathbb{C}} \int_{\mathbb{C}} p(y_k | x_k) p(x_k) dx_k \log \left( \int_{\mathbb{C}} p(y_k | x_k) p(x_k) dx_k \right) dy_k \\ &= -\int_0^{\infty} \left[ \int_0^{\infty} \frac{2|y|}{\sigma_h^2 |x|^2 + \sigma_n^2} e^{-\frac{|y|^2}{\sigma_h^2 |x|^2 + \sigma_n^2}} \frac{2|x|}{\sigma_x^2} e^{-\frac{|x|^2}{\sigma_x^2}} d|x| \right] \\ &\quad \times \log \left( \int_0^{\infty} \frac{2|y|}{\sigma_h^2 |x|^2 + \sigma_n^2} e^{-\frac{|y|^2}{\sigma_h^2 |x|^2 + \sigma_n^2}} \frac{2|x|}{\sigma_x^2} e^{-\frac{|x|^2}{\sigma_x^2}} d|x| \right) d|y| \\ &\quad + \log(2\pi) - \frac{\gamma}{2} + \frac{1}{2} \int_{z=0}^{\infty} \log(\sigma_h^2 \sigma_x^2 z + \sigma_n^2) e^{-z} dz \end{aligned} \quad (168)$$

$$= h'_{U_2}(\mathbf{y}) \quad (169)$$

where  $\gamma \approx 0.57721$  is the Euler constant. To the best of our knowledge, the first integral in (168) cannot be calculated analytically. However, it can be evaluated numerically using Hermite polynomials and Simpson's rule, see [21], [31], [32], or by Monte Carlo integration.

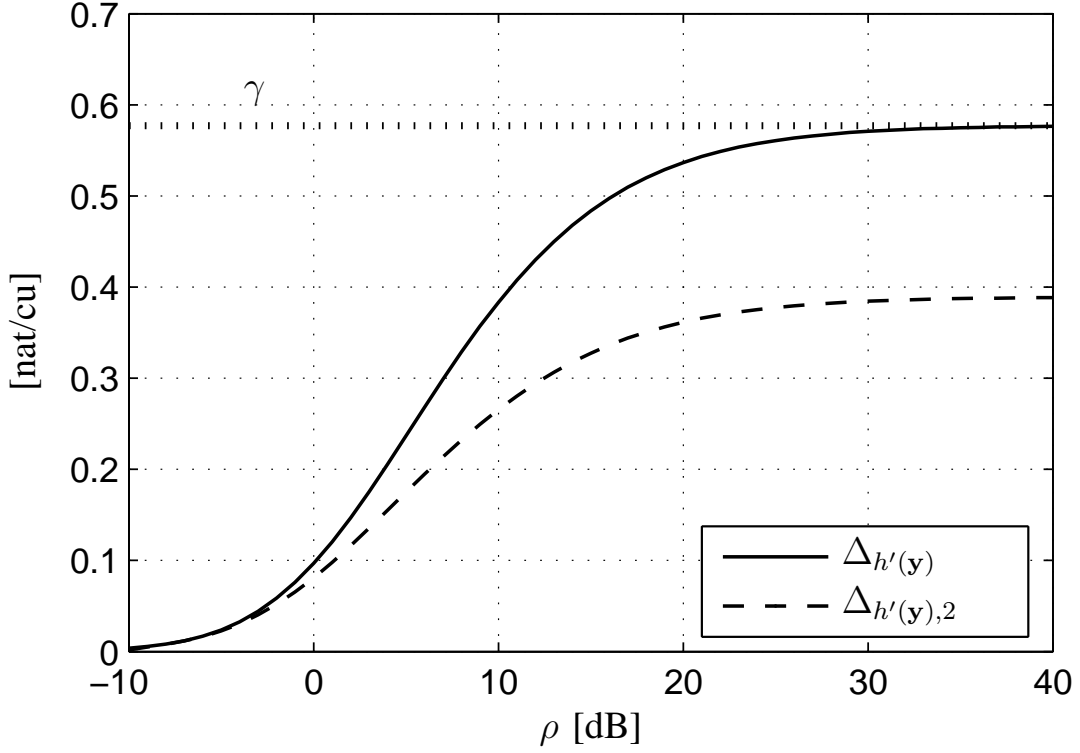


Fig. 7. Comparison of  $\Delta_{h'(\mathbf{y}),2}$  with  $\Delta_{h'(\mathbf{y})}$

For the evaluation of the tightness of  $h'_{U_2}(\mathbf{y})$ , in Fig. 7 the difference

$$\Delta_{h'(\mathbf{y}),2} = h'_{U_2}(\mathbf{y}) - h'_L(\mathbf{y}) \quad (170)$$

is shown in comparison to the difference  $\Delta_{h'(\mathbf{y})}$  given in (88). Obviously, the upper bound  $h'_{U_2}(\mathbf{y})$  is tighter than the upper bound  $h'_U(\mathbf{y})$  given in (36).

#### APPENDIX C

##### SUFFICIENT CONDITIONS FOR $\alpha_{\text{OPT}} = 1$ IN (119)

In this appendix, we give conditions on the parameters  $f_d$ ,  $\rho$ , and  $\beta$  such that  $\alpha_{\text{opt}} = 1$  in (119), i.e., the upper bound in (117) is maximized by choosing the maximum average power  $\sigma_x^2$ . Therefore, we have to evaluate for which parameter choice the following inequality holds:

$$\begin{aligned} \left( \frac{2f_d}{\beta} \log \left( \frac{\rho\beta}{2f_d} + 1 \right) \right)^{-1} - \frac{1}{\rho} &\geq 1 \\ \Leftrightarrow \frac{2f_d}{\beta} \log \left( \frac{\rho\beta}{2f_d} + 1 \right) &\leq \frac{\rho}{1+\rho}. \end{aligned} \quad (171)$$

The following calculations are closely related to a corresponding problem in [8, Appendix C]. We divide the evaluation into the two cases  $\rho > 1$  and  $\rho \leq 1$ .

For  $\rho > 1$ , the RHS of (171) can be lower-bounded by

$$\frac{\rho}{1+\rho} \geq \frac{1}{2} \quad (172)$$



yielding the following sufficient condition for (171) to hold:

$$\begin{aligned} \frac{2f_d}{\beta} \log \left( \frac{\rho\beta}{2f_d} + 1 \right) &\leq \frac{1}{2} \\ \Leftrightarrow \rho &\leq \frac{2f_d}{\beta} \left[ \exp \left( \frac{1}{2} \frac{\beta}{2f_d} \right) - 1 \right]. \end{aligned} \quad (173)$$

Thus,  $\alpha_{\text{opt}} = 1$  holds if

$$1 < \rho \leq \frac{2f_d}{\beta} \left[ \exp \left( \frac{1}{2} \frac{\beta}{2f_d} \right) - 1 \right]. \quad (174)$$

Now, we discuss the case  $\rho \leq 1$ . Using the inequality  $\frac{1}{x} \log(x+1) \leq \frac{1}{\sqrt{x+1}}$  for  $x \geq 0$ , for  $\rho \leq 1$  the LHS of (171) can be upper-bounded by

$$\frac{2f_d}{\beta} \log \left( \frac{\rho\beta}{2f_d} + 1 \right) \leq \frac{\rho}{\sqrt{\frac{\rho\beta}{2f_d} + 1}}. \quad (175)$$

Based on (175), inequality (171) holds if the following sufficient condition is fulfilled:

$$\begin{aligned} \frac{\rho}{\sqrt{\frac{\rho\beta}{2f_d} + 1}} &\leq \frac{\rho}{1 + \rho} \\ \Leftrightarrow 2f_d &\leq \frac{\beta}{\rho + 2} \end{aligned} \quad (176)$$

so that we get the second condition

$$2f_d \leq \frac{\beta}{\rho + 2} \text{ for } \rho \leq 1. \quad (177)$$

Thus, if (174) or (177) is fulfilled, (119) yields  $\alpha_{\text{opt}} = 1$ .

#### APPENDIX D CONVEXITY OF (149)

To prove that (149) is convex with respect to the individual diagonal elements of  $\mathbf{Z}$ , we rewrite the prediction error variance  $\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1}) = \sigma_{e_{\text{pred}}}^2(\mathbf{z})$  as follows:

$$\begin{aligned} \sigma_{e_{\text{pred}}}^2(\mathbf{z}) &= \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H (\mathbf{R}_h + \sigma_n^2 \mathbf{Z}^{-1})^{-1} \mathbf{r}_{h,\text{pred}} \\ &\stackrel{(a)}{=} \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \left( \mathbf{R}_h^{-1} - \mathbf{R}_h^{-1} \left( \frac{\mathbf{Z}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \right) \mathbf{r}_{h,\text{pred}} \\ &\stackrel{(b)}{=} \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \left( \mathbf{R}_h^{-1} - \mathbf{R}_h^{-1} \left( \frac{z_i \mathbf{V}_i + \mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \right) \mathbf{r}_{h,\text{pred}} \\ &= \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \left( \mathbf{R}_h^{-1} - \mathbf{R}_h^{-1} \left[ \left( \frac{1}{\sigma_n^2} \mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1} \right) \left\{ \left( \frac{1}{\sigma_n^2} \mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1} \right)^{-1} \frac{z_i}{\sigma_n^2} \mathbf{V}_i + \mathbf{I} \right\} \right]^{-1} \mathbf{R}_h^{-1} \right) \mathbf{r}_{h,\text{pred}} \\ &= \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \left( \mathbf{R}_h^{-1} - \mathbf{R}_h^{-1} \left\{ \left( \frac{1}{\sigma_n^2} \mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1} \right)^{-1} \frac{z_i}{\sigma_n^2} \mathbf{V}_i + \mathbf{I} \right\}^{-1} \left( \frac{1}{\sigma_n^2} \mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \right) \mathbf{r}_{h,\text{pred}} \\ &\stackrel{(c)}{=} \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \left( \mathbf{R}_h^{-1} - \mathbf{R}_h^{-1} \left\{ \mathbf{I} - \frac{z_i}{1 + z_i \lambda_{\max}} \left( \frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \frac{\mathbf{V}_i}{\sigma_n^2} \right\} \left( \frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \right) \mathbf{r}_{h,\text{pred}} \end{aligned}$$

$$\begin{aligned}
&= \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \left( \mathbf{R}_h^{-1} - \mathbf{R}_h^{-1} \left( \frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \right) \mathbf{r}_{h,\text{pred}} \\
&\quad - \frac{z_i \mathbf{r}_{h,\text{pred}}^H \mathbf{R}_h^{-1} \left( \frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \frac{\mathbf{V}_i}{\sigma_n^2} \left( \frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \mathbf{r}_{h,\text{pred}}}{1 + z_i \lambda_{\max}} \\
&\stackrel{(d)}{=} \sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{z_i \cdot a}{1 + z_i \lambda_{\max}}
\end{aligned} \tag{178}$$

where for (a) we have used the matrix inversion lemma, and for (b) we have separated the diagonal matrix  $\mathbf{Z}$  as follows:

$$\mathbf{Z} = \mathbf{Z}_{\setminus i} + z_i \mathbf{V}_i \tag{179}$$

where  $\mathbf{Z}_{\setminus i}$  corresponds to  $\mathbf{Z}$  except that the  $i$ -th diagonal element is set to 0,  $\mathbf{V}_i$  is a matrix with all elements zero except of the  $i$ -th diagonal element being equal to 1, and  $z_i$  is the  $i$ -th diagonal element of the matrix  $\mathbf{Z}$ . In addition, for (c) we have used the Sherman-Morrison formula and  $\lambda_{\max}$  is the non-zero eigenvalue of the rank one matrix

$$\mathbf{B} = \left( \frac{1}{\sigma_n^2} \mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1} \right)^{-1} \frac{1}{\sigma_n^2} \mathbf{V}_i. \tag{180}$$

Furthermore, for (d) we substituted  $\sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i})$  for

$$\sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) = \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \left( \mathbf{R}_h^{-1} - \mathbf{R}_h^{-1} \left( \frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \right) \mathbf{r}_{h,\text{pred}}$$

which is the prediction error variance if the observation at the  $i$ -th time instant is not used for channel prediction. Additionally, for (d) we have also used the substitution

$$\begin{aligned}
a &= \mathbf{r}_{h,\text{pred}}^H \mathbf{R}_h^{-1} \left( \frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \frac{\mathbf{V}_i}{\sigma_n^2} \left( \frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \mathbf{r}_{h,\text{pred}} \\
&\geq 0
\end{aligned} \tag{181}$$

where the nonnegativity follows as  $\mathbf{V}_i$  is positive semidefinite.

Thus, with (178) we have found a separation of the channel prediction error variance  $\sigma_{e_{\text{pred}}}^2(\mathbf{z})$  into the term  $\sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i})$  being independent of  $z_i$ , and an additional term, which depends on  $z_i$ . Note that  $a$  and  $\lambda_{\max}$  in the second term on the RHS of (178) are independent of  $z_i$  and that the element  $i$  is an arbitrarily chosen element. I.e., we can use this separation for each diagonal element of the matrix  $\mathbf{Z}$ .

By substituting the RHS of (178) into (149) we get

$$\log \left( 1 + \frac{|x_N|^2}{\sigma_n^2} \left( \sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{z_i \cdot a}{1 + z_i \lambda_{\max}} \right) \right) = K. \tag{182}$$

Recall that we want to show the convexity of (182) with respect to the element  $z_i$ . Therefore, we calculate its second derivative with respect to  $z_i$  which is given by

$$\frac{\partial^2 K}{(\partial z_i)^2} = \frac{\frac{|x_N|^2}{\sigma_n^2} \frac{a 2 \lambda_{\max} (1 + z_i \lambda_{\max})}{(1 + z_i \lambda_{\max})^4} \left\{ 1 + \frac{|x_N|^2}{\sigma_n^2} \left( \sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{a(z_i + \frac{1}{2\lambda_{\max}})}{1 + z_i \lambda_{\max}} \right) \right\}}{\left( 1 + \frac{|x_N|^2}{\sigma_n^2} \left( \sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{a z_i}{1 + z_i \lambda_{\max}} \right) \right)^2}$$

and we will show that it is nonnegative, i.e.,

$$\frac{\partial^2 K}{(\partial z_i)^2} \geq 0. \tag{183}$$

Therefore, first we show that  $\lambda_{\max}$  is nonnegative. This can be done based on the definition of the eigenvalues of the matrix  $\mathbf{B}$

$$\begin{aligned}\mathbf{B}\mathbf{u} &= \left( \frac{1}{\sigma_n^2} \mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1} \right)^{-1} \frac{1}{\sigma_n^2} \mathbf{V}_i \mathbf{u} = \lambda_{\max} \mathbf{u} \\ \Rightarrow \frac{1}{\sigma_n^2} \mathbf{u}^H \mathbf{V}_i \mathbf{u} &= \lambda_{\max} \mathbf{u}^H \left( \frac{1}{\sigma_n^2} \mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1} \right) \mathbf{u} \\ &\stackrel{(a)}{\Rightarrow} \lambda_{\max} \geq 0\end{aligned}$$

where (a) follows from the fact that the eigenvalues of  $\left( \frac{1}{\sigma_n^2} \mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1} \right)$  are nonnegative, as  $\mathbf{R}_h$  is positive definite and the diagonal entries of the diagonal matrix  $\mathbf{Z}_{\setminus i}$  are also nonnegative. In addition,  $\mathbf{V}_i$  is also positive semidefinite.

With  $\lambda_{\max}$ ,  $z_i$ , and  $a$  being nonnegative, for the proof of (183), it rests to show that

$$\sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{a}{1 + z_i \lambda_{\max}} \left( z_i + \frac{1}{2\lambda_{\max}} \right) \geq 0. \quad (184)$$

To prove this inequality, we calculate the derivative of the LHS of (184) with respect to  $z_i$ , which is given by

$$\frac{\partial}{\partial z_i} \left\{ \sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{a \left( z_i + \frac{1}{2\lambda_{\max}} \right)}{1 + z_i \lambda_{\max}} \right\} = -\frac{a}{2(1 + z_i \lambda_{\max})^2} \leq 0 \quad (185)$$

where for the last inequality we have used (181). I.e., the LHS of (184) monotonically decreases in  $z_i$ . Furthermore, for  $z_i \rightarrow \infty$  the LHS of (184) becomes

$$\lim_{z_i \rightarrow \infty} \left\{ \sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{a \left( z_i + \frac{1}{2\lambda_{\max}} \right)}{1 + z_i \lambda_{\max}} \right\} \stackrel{(a)}{=} \lim_{z_i \rightarrow \infty} \sigma_{e_{\text{pred}}}^2(\mathbf{z}) \stackrel{(b)}{\geq} 0 \quad (186)$$

where (a) follows due to (178), and where (b) holds as the prediction error variance must be nonnegative. As the LHS of (184) is monotonically decreasing in  $z_i$  and as its limit for  $z_i \rightarrow \infty$  is nonnegative, (184) must hold.

Thus, with (184) inequality (183) holds and, thus, (182) is convex in  $z_i$ . As the element  $i$  has been chosen arbitrarily, in conclusion, we have shown that (149) is convex in each  $z_i$  for  $i = 1, \dots, N-1$ .

## REFERENCES

- [1] V. Sethuraman, L. Wang, B. Hajek, and A. Lapidath, "Low-SNR capacity of noncoherent fading channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1555–1574, Apr. 2009.
- [2] V. Sethuraman, B. Hajek, and K. Narayanan, "Capacity bounds for noncoherent fading channels with a peak constraint," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, Sep. 2005, pp. 515–519.
- [3] A. Lapidath, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.
- [4] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.
- [5] J. Baltersee, G. Fock, and H. Meyr, "An information theoretic foundation of synchronized detection," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2115–2123, Dec. 2001.
- [6] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [7] R. H. Etkin and D. N. C. Tse, "Degrees of freedom in some underspread MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1576–1608, Apr. 2006.
- [8] G. Durisi, U. G. Schuster, H. Bölcskei, and S. Shamai (Shitz), "Noncoherent capacity of underspread fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 367–395, Jan. 2010.
- [9] T. Koch and A. Lapidath, "On multipath fading channels at high SNR," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 5945–5957, Dec. 2010.

- [10] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need "perfect side information" be?" *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
- [11] F. Rusek, A. Lozano, and N. Jindal, "Mutual information of IID complex Gaussian signals on block Rayleigh-faded channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, U.S.A., Jun. 2010, pp. 300–304.
- [12] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.
- [13] R.-R. Chen, B. Hajek, R. Koetter, and U. Madhow, "On fixed input distributions for noncoherent communication over high-SNR Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3390–3396, Dec. 2004.
- [14] J. Doob, *Stochastic Processes*. New York: Wiley, 1990.
- [15] W. C. Jakes, Ed., *Microwave Mobile Communications*. New York, NY, U.S.A.: John Wiley & Sons Inc, 1975.
- [16] X. Deng and A. M. Haimovich, "Information rates of time varying Rayleigh fading channels," in *Proc. IEEE International Conference on Communications (ICC)*, vol. 1, Paris, France, Jun. 2004, pp. 573–577.
- [17] M. C. Valenti and B. D. Woerner, "Iterative channel estimation and decoding of pilot symbol assisted Turbo codes over flat-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1697–1705, Sep. 2001.
- [18] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.
- [19] V. Sethuraman and B. Hajek, "Capacity per unit energy of fading channels with a peak constraint," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3102–3120, Sep. 2005.
- [20] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [21] R. R. Perera, T. S. Pollock, and T. D. Abhayapala, "Performance of Gaussian distributed input in non-coherent Rayleigh fading MIMO channels," in *Proc. IEEE International Conference on Communications and Signal Processing (ICICS2005)*, Bangkok, Thailand, Dec. 2005, pp. 791–795.
- [22] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [23] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.
- [24] U. Grenander and G. Szegő, *Toeplitz Forms and Their Applications*. Berkeley, CA, U.S.A.: Univ. Calif. Press, 1958.
- [25] D. Guo, S. Shamai (Shitz), and S. Verdú, "Proof of entropy power inequalities via MMSE," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Seattle, WA, Jul. 2006, pp. 1011–1015.
- [26] J. Chen and V. V. Veeravalli, "Capacity results for block-stationary Gaussian fading channels with a peak power constraint," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4498–4520, Dec. 2007.
- [27] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.
- [28] T. Cover and J. Thomas, *Elements of Information Theory, 2nd edition*. New York: Wiley & Sons, 2006.
- [29] M. Dörpinghaus, A. Ispas, G. Ascheid, and H. Meyr, "On the gain of joint processing of pilot and data symbols in stationary Rayleigh fading channels," in *Proc. of the 2010 International Zurich Seminar on Communications*, Zurich, Switzerland, Mar. 2010, pp. 74–77.
- [30] L. Schmitt, H. Meyr, and D. Zhang, "Systematic design of iterative ML receivers for flat fading channels," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 1897–1901, Jul. 2010.
- [31] R. R. Perera, "On capacity of fading channels with no channel state information," Ph.D. dissertation, The Australian National University, June 2006.
- [32] N. M. Steen, G. D. Byrne, and E. M. Gelbard, "Gaussian quadrature for the integrals," *Mathematics of Computation*, vol. 23, no. 107, pp. 661–671, Jul. 1969.